(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 July 2002 (18.07.2002)

PCT

(10) International Publication Number
WO 02/055657 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/48561

(22) International Filing Date:
20 December 2001 (20.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/256,983 21 December 2000 (21.12.2000) US

(71) Applicant: **MYRIAD GENETICS, INC.** [US/US]; 320 Wakara Way, Salt Lake City, UT 84108 (US).

(72) Inventors: **CIMBORA, Daniel, M.**; 2134 S. Wyoming Street, Salt Lake City, UT 84109 (US). **HEICHMAN, Karen**; 173 South 1200 East, Salt Lake City, UT 84102 (US). **BARTEL, Paul, L.**; 1461 Kensington Avenue, Salt Lake City, UT 84105 (US).

(74) Agents: **IHNEN, Jeffrey, L.** et al.; Rothwell, Figg, Ernst & Manbeck, P.C., 1425 K Street N.W., Suite 800, Washington, DC 20005 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
*without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/055657 A2

(54) Title: PROTEIN-PROTEIN INTERACTIONS

(57) Abstract: The present invention relates to the discovery of novel protein-protein interactions that are involved in mammalian physiological pathways, including physiological disorders or diseases. Examples of physiological disorders and diseases include non-insulin dependent diabetes mellitus (NIDDM), neurodegenerative disorders, such as Alzheimer's Disease (AD), and the like. Thus, the present invention is directed to complexes of these proteins and/or their fragments, antibodies to the complexes, diagnosis of physiological generative disorders (including diagnosis of a predisposition to and diagnosis of the existence of the disorder), drug screening for agents which modulate the interaction of proteins described herein, and identification of additional proteins in the pathway common to the proteins described herein.

# PROTEIN-PROTEIN INTERACTIONS

## BACKGROUND OF THE INVENTION

[0001] The present invention relates to the discovery of novel protein-protein interactions that are involved in mammalian physiological pathways, including physiological disorders or diseases. Examples of physiological disorders and diseases include non-insulin dependent diabetes mellitus (NIDDM), neurodegenerative disorders, such as Alzheimer's Disease (AD), and the like. Thus, the present invention is directed to complexes of these proteins and/or their fragments, antibodies to the complexes, diagnosis of physiological generative disorders (including diagnosis of a predisposition to and diagnosis of the existence of the disorder), drug screening for agents which modulate the interaction of proteins described herein, and identification of additional proteins in the pathway common to the proteins described herein.

[0002] The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended Bibliography.

[0003] Many processes in biology, including transcription, translation and metabolic or signal transduction pathways, are mediated by non-covalently associated protein complexes. The formation of protein-protein complexes or protein-DNA complexes produce the most efficient chemical machinery. Much of modern biological research is concerned with identifying proteins involved in cellular processes, determining their functions, and how, when and where they interact with other proteins involved in specific pathways. Further, with rapid advances in genome sequencing, there is a need to define protein linkage maps, i.e., detailed inventories of protein interactions that make up functional assemblies of proteins or protein complexes or that make up physiological pathways.

[0004] Recent advances in human genomics research has led to rapid progress in the identification of novel genes. In applications to biological and pharmaceutical research, there is a need to determine functions of gene products. A first step in defining the function of a novel gene is to determine its interactions with other gene products in appropriate context. That is, since proteins make specific interactions with other proteins or other biopolymers as part of functional assemblies or physiological pathways, an appropriate way to examine function of a gene is to

determine its physical relationship with other genes. Several systems exist for identifying protein interactions and hence relationships between genes.

[0005] There continues to be a need in the art for the discovery of additional protein-protein interactions involved in mammalian physiological pathways. There continues to be a need in the art also to identify the protein-protein interactions that are involved in mammalian physiological disorders and diseases, and to thus identify drug targets.

## SUMMARY OF THE INVENTION

[0006] The present invention relates to the discovery of protein-protein interactions that are involved in mammalian physiological pathways, including physiological disorders or diseases, and to the use of this discovery. The identification of the interacting proteins described herein provide new targets for the identification of useful pharmaceuticals, new targets for diagnostic tools in the identification of individuals at risk, sequences for production of transformed cell lines, cellular models and animal models, and new bases for therapeutic intervention in such physiological pathways

[0007] Thus, one aspect of the present invention is protein complexes. The protein complexes are a complex of (a) two interacting proteins, (b) a first interacting protein and a fragment of a second interacting protein, (c) a fragment of a first interacting protein and a second interacting protein, or (d) a fragment of a first interacting protein and a fragment of a second interacting protein. The fragments of the interacting proteins include those parts of the proteins, which interact to form a complex. This aspect of the invention includes the detection of protein interactions and the production of proteins by recombinant techniques. The latter embodiment also includes cloned sequences, vectors, transfected or transformed host cells and transgenic animals.

[0008] A second aspect of the present invention is an antibody that is immunoreactive with the above complex. The antibody may be a polyclonal antibody or a monoclonal antibody. While the antibody is immunoreactive with the complex, it is not immunoreactive with the component parts of the complex. That is, the antibody is not immunoreactive with a first interactive protein, a fragment of a first interacting protein, a second interacting protein or a fragment of a second interacting protein. Such antibodies can be used to detect the presence or absence of the protein complexes.

[0009] A third aspect of the present invention is a method for diagnosing a predisposition for physiological disorders or diseases in a human or other animal. The diagnosis of such disorders

includes a diagnosis of a predisposition to the disorders and a diagnosis for the existence of the disorders. In accordance with this method, the ability of a first interacting protein or fragment thereof to form a complex with a second interacting protein or a fragment thereof is assayed, or the genes encoding interacting proteins are screened for mutations in interacting portions of the protein molecules. The inability of a first interacting protein or fragment thereof to form a complex, or the presence of mutations in a gene within the interacting domain, is indicative of a predisposition to, or existence of a disorder. In accordance with one embodiment of the invention, the ability to form a complex is assayed in a two-hybrid assay. In a first aspect of this embodiment, the ability to form a complex is assayed by a yeast two-hybrid assay. In a second aspect, the ability to form a complex is assayed by a mammalian two-hybrid assay. In a second embodiment, the ability to form a complex is assayed by measuring *in vitro* a complex formed by combining said first protein and said second protein. In one aspect the proteins are isolated from a human or other animal. In a third embodiment, the ability to form a complex is assayed by measuring the binding of an antibody, which is specific for the complex. In a fourth embodiment, the ability to form a complex is assayed by measuring the binding of an antibody that is specific for the complex with a tissue extract from a human or other animal. In a fifth embodiment, coding sequences of the interacting proteins described herein are screened for mutations.

[0010] A fourth aspect of the present invention is a method for screening for drug candidates which are capable of modulating the interaction of a first interacting protein and a second interacting protein. In this method, the amount of the complex formed in the presence of a drug is compared with the amount of the complex formed in the absence of the drug. If the amount of complex formed in the presence of the drug is greater than or less than the amount of complex formed in the absence of the drug, the drug is a candidate for modulating the interaction of the first and second interacting proteins. The drug promotes the interaction if the complex formed in the presence of the drug is greater and inhibits (or disrupts) the interaction if the complex formed in the presence of the drug is less. The drug may affect the interaction directly, i.e., by modulating the binding of the two proteins, or indirectly, e.g., by modulating the expression of one or both of the proteins.

[0011] A fifth aspect of the present invention is a model for such physiological pathways, disorders or diseases. The model may be a cellular model or an animal model, as further described herein. In accordance with one embodiment of the invention, an animal model is prepared by creating transgenic or "knock-out" animals. The knock-out may be a total knock-out, i.e., the desired gene is deleted, or a conditional knock-out, i.e., the gene is active until it is knocked out at

a determined time. In a second embodiment, a cell line is derived from such animals for use as a model. In a third embodiment, an animal model is prepared in which the biological activity of a protein complex of the present invention has been altered. In one aspect, the biological activity is altered by disrupting the formation of the protein complex, such as by the binding of an antibody or small molecule to one of the proteins which prevents the formation of the protein complex. In a second aspect, the biological activity of a protein complex is altered by disrupting the action of the complex, such as by the binding of an antibody or small molecule to the protein complex which interferes with the action of the protein complex as described herein. In a fourth embodiment, a cell model is prepared by altering the genome of the cells in a cell line. In one aspect, the genome of the cells is modified to produce at least one protein complex described herein. In a second aspect, the genome of the cells is modified to eliminate at least one protein of the protein complexes described herein.

[0012] A sixth aspect of the present invention are nucleic acids coding for novel proteins discovered in accordance with the present invention and the corresponding proteins and antibodies.

[0013] A seventh aspect of the present invention is a method of screening for drug candidates useful for treating a physiological disorder. In this embodiment, drugs are screened on the basis of the association of a protein with a particular physiological disorder. This association is established in accordance with the present invention by identifying a relationship of the protein with a particular physiological disorder. The drugs are screened by comparing the activity of the protein in the presence and absence of the drug. If a difference in activity is found, then the drug is a drug candidate for the physiological disorder. The activity of the protein can be assayed *in vitro* or *in vivo* using conventional techniques, including transgenic animals and cell lines of the present invention.

## DETAILED DESCRIPTION OF THE INVENTION

[0014] The present invention is the discovery of novel interactions between proteins described herein. The genes coding for some of these proteins may have been cloned previously, but their potential interaction in a physiological pathway or with a particular protein was unknown. Alternatively, the genes coding for some of these proteins have not been cloned previously and represent novel genes. These proteins are identified using the yeast two-hybrid method and searching a human total brain library, as more fully described below.

[0015] According to the present invention, new protein-protein interactions have been discovered. The discovery of these interactions has identified several protein complexes for each protein-protein interaction. The protein complexes for these interactions are set forth below in Tables 1-10, which also identifies the new protein-protein interactions of the present invention.

TABLE 1

Protein Complexes LXR-alpha/Utrophin Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and utrophin

A fragment of LXR-alpha and utrophin

LXR-alpha and a fragment of utrophin

A fragment of LXR-alpha and a fragment of utrophin

TABLE 2

Protein Complexes LXR-alpha/zyxin Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and zyxin

A fragment of LXR-alpha and zyxin

LXR-alpha and a fragment of zyxin

A fragment of LXR-alpha and a fragment of zyxin

TABLE 3

Protein Complexes LXR-alpha/LIMS1 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and LIMS1

A fragment of LXR-alpha and LIMS1

LXR-alpha and a fragment of LIMS1

A fragment of LXR-alpha and a fragment of LIMS1

TABLE 4

Protein Complexes LXR-alpha/PN7771 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and PN7771

A fragment of LXR-alpha and PN7771

LXR-alpha and a fragment of PN7771

A fragment of LXR-alpha and a fragment of PN7771

TABLE 5

Protein Complexes LXR-alpha/Homer-3 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and Homer-3

A fragment of LXR-alpha and Homer-3

LXR-alpha and a fragment of Homer-3

A fragment of LXR-alpha and a fragment of Homer-3

TABLE 6

Protein Complexes LXR-alpha/RACK1 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and RACK1

A fragment of LXR-alpha and RACK1

LXR-alpha and a fragment of RACK1

A fragment of LXR-alpha and a fragment of RACK1

TABLE 7

Protein Complexes LXR-alpha/EIF3S1 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and EIF3S1

A fragment of LXR-alpha and EIF3S1

LXR-alpha and a fragment of EIF3S1

A fragment of LXR-alpha and a fragment of EIF3S1

TABLE 8

Protein Complexes LXR-alpha/PSMD11 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and PSMD11

A fragment of LXR-alpha and PSMD11

LXR-alpha and a fragment of PSMD11

A fragment of LXR-alpha and a fragment of PSMD11

TABLE 9

Protein Complexes LXR-alpha/KIAA0610 Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and KIAA0610

A fragment of LXR-alpha and KIAA0610

LXR-alpha and a fragment of KIAA0610

A fragment of LXR-alpha and a fragment of KIAA0610

TABLE 10

Protein Complexes LXR-alpha/CIR Interaction

Oxysterol liver X receptor alpha (LXR-alpha) and CIR

A fragment of LXR-alpha and CIR

LXR-alpha and a fragment of CIR

A fragment of LXR-alpha and a fragment of CIR

[0016] The involvement of above interactions in particular pathways is as follows.

[0017] Many cellular proteins exert their function by interacting with other proteins in the cell. Examples of this are found in the formation of multiprotein complexes and the association of enzymes with their substrates. It is widely believed that a great deal of information can be gained by understanding individual protein-protein interactions, and that this is useful in identifying complex networks of interacting proteins that participate in the workings of normal cellular functions. Ultimately, the knowledge gained by characterizing these networks can lead to valuable insight into the causes of human diseases and can eventually lead to the development of therapeutic strategies. The yeast two-hybrid assay is a powerful tool for determining protein-protein interactions and it has been successfully used for studying human disease pathways. In one variation of this technique, a protein of interest (or a portion of that protein) is expressed in a

population of yeast cells that collectively contain all protein sequences. Yeast cells that possess protein sequences that interact with the protein of interest are then genetically selected, and the identity of those interacting proteins are determined by DNA sequencing. Thus, proteins that can be demonstrated to interact with a protein known to be involved in a human disease are therefore also implicated in that disease. Proteins identified in the first round of two-hybrid screening can be subsequently used in a second round of two-hybrid screening, allowing the identification of multiple proteins in the complex network of interactions in a disease pathway.

[0018] Nuclear hormone receptors play important roles in development, reproduction, and physiology by altering gene transcription in response to hormonal signals (Whitfield et al., 1999; Klein-Hitpass et al., 1998). Misregulation of hormone receptor signaling pathways is responsible for a variety of diseases. For example, aldosterone and its receptor (the mineralocorticoid receptor, MCR) are involved in hypertension and congestive heart failure (Duprez et al., 2000), and it has recently been shown that a missense mutation in MCR that alters its ligand specificity is responsible for pregnancy-exacerbated hypertension (Geller et al., 2000). Likewise, glucocorticoids and the glucocorticoid receptor (GR) have been implicated in chronic inflammation and arthritis (Banres, 1998), and the oxysterol liver receptor (LXR), farnesoid X receptor (FXR), and other nuclear receptors are involved in cholesterol homeostasis and atherogenesis (Schroepfer, 2000; Haynes et al., 2000; Brown and Jessup, 1999)

[0019] Collectively, the nuclear receptor superfamily is responsive to a wide variety of ligands. Nuclear hormone receptors share several important structural features, including a variable N-terminal region, a conserved central DNA-binding domain, a variable hinge region, and a conserved C-terminal ligand-binding domain (Moras and Gronemeyer, 1998; Mangelsdorf et al., 1995). Despite this conserved structural organization, interactions between ligands and receptors are remarkably specific. Hormone binding results in conformational changes in the receptor, allowing binding to specific DNA sequences (hormone response elements, HREs) in target gene promoters resulting in changes in target gene transcription. Interaction of nuclear hormone receptors with accessory proteins determines whether the receptor activates or represses transcription. Receptors can recruit coactivators that remodel chromatin and stabilize the RNA polymerase machinery, or alternatively can interact with factors that condense chromatin structure and inactivate gene expression (Wolffe et al., 1997). Furthermore, binding of a nuclear hormone receptor to other cellular proteins can alter the subcellular localization of the receptor and control its ability to bind hormone and HREs (DeFranco et al., 1998). Clearly, identification of factors with which nuclear

hormone receptors interact is vital to understanding the process by which hormonal signals are transduced into transcriptional responses. In addition, identification of receptor-interacting proteins will increase the repertoire of potential targets for therapeutic intervention in the treatment of diseases due to defects involving nuclear hormone signaling.

[0020] The oxysterol liver X receptor alpha (LXRa) was used in yeast two-hybrid searches to identify novel protein-protein interactions. Here, we describe ten new interactors for LXRa. The first four interactors are involved in cell adhesion and cellular architecture. The first of these is the actin-binding protein utrophin. Utrophin is an autosomal gene that is similar to dystrophin, the gene famous for its role in Duchenne's muscular dystrophy. Unlike, dystrophin, however, utrophin appears to be expressed in a wide variety of adult tissues. Dystrophin and the dystrophin-related proteins contain spectrin repeats and likely play a role in anchoring the cytoskeleton to the plasma membrane by their actin-binding activities. The second interactor is the adhesion plaque protein zyxin, also involved in anchoring the cytoskeleton. Zyxin is a phosphoprotein that contains three LIM domains and two proline-rich regions. The interaction of LXRa with zyxin is reminiscent of the interaction we have identified between the farnesoid X-activated receptor and the LIM domain cytoskeletal protein Paxillin. Third, LXRa interacts with the novel protein PN7771, which is highly related (greater than 90% amino acid identity) to Ninein. Ninein is a centrosome-associated protein that interacts with human glycogen synthase kinase 3beta (GSK-3beta) (Hong et al., 2000), is localized to the pericentriolar matrix of the centrosome, and reacts with centrosomal autoantibody sera (Mack et al., 1998). PN7771 contains predicted calcium-binding EF hand motifs, a potential nuclear localization signal, a basic region-leucine zipper motif, a spectrin repeat, coiled-coil motifs, and Glu- and Gln-rich regions. Taken together, these interactions suggest that LXRa may be involved in cellular signaling events in response to cellular adhesion or other extracellular stimuli, and that the trans-activating ability of LXRa may be regulated by its interaction with these proteins.

[0021] Several LXRa interactors are involved in signal transduction pathways. The first is the neuronal immediate early protein homer-3. Homer proteins bind to the C-terminal tails of metabotropic glutamate receptors and play a role in their targeting and regulation; the metabotropic glutamate receptors, in turn, participate in the influx of intracellular calcium. Since LXR-alpha binds to homer-3, it is possible that LXRa may also be involved in calcium release. Alternatively, LXR-alpha could be modulated in some way by homer-3 in a manner analogous to the way in which the metabotropic glutamate receptors are regulated. The second protein, RACK1 (receptor of activated protein kinase C 1), is a WD repeat-containing protein that functions as an intracellular

receptor to localize PKC to the cytoskeleton. The interaction between RACK1 and LXR-alpha suggests that LXR-alpha may be capable of localizing to the cytoskeleton via its association with RACK1. The next interactor is the LIM-domain protein LIMS1. LIMS1 has been implicated in integrin-linked kinase signaling, and it has been shown to interact with the SH3 and SH2 domain-containing adaptor protein NCK2 (Tu et al. 1998). Taken together, these findings suggest the involvement of LXRa in a variety of signal transduction pathways; whether LXRa activity is regulated by interaction with these proteins, or vice versa, remains to be determined.

[0022] Two LXRa interactors involved in protein metabolism were identified: the proteasome subunit PSMD11, which is involved in protein turnover, and the translation initiation factor EIF3S1, which is involved in protein synthesis. The interaction of these proteins with LXRa suggests that nuclear hormone receptors may be involved directly with protein production and stability, in addition to transcriptional regulation.

[0023] An interaction between LXRa and the potential transmembrane protein KIAA0610 was identified. KIAA0610 is hypothetical protein fragment 686 amino acids in length. Predicted structural motifs include four possible transmembrane domains and a coiled-coil domain. KIAA0610 displays weak homology (~24% amino acid identity over 360-430 residues) to *Drosophila* and *C. elegans* proteins of unknown function. EST analysis suggests expression in a variety of tissues.

[0024] Finally, and interaction between LXRa and the transcription factor CIR was identified. CIR has been demonstrated to interact with the CBF1 transcription factor as well as histone deacetylase HD2 and Sin3-associated protein 30kD (Hsieh et al., 1999). It has been proposed that CIR acts as a linker between CBF1 and the histone deacetylase complex. Similarly, the interaction of LXRa with CIR suggests CIR may link LXRa with the histone deacetylase machinery. In support of a functional role between nuclear receptors and CIR, we have also identified an interaction between CIR and the estrogen receptor ER-beta.

[0025] The proteins disclosed in the present invention were found to interact with their corresponding proteins in the yeast two-hybrid system. Because of the involvement of the corresponding proteins in the physiological pathways disclosed herein, the proteins disclosed herein also participate in the same physiological pathways. Therefore, the present invention provides a list of uses of these proteins and DNA encoding these proteins for the development of diagnostic and therapeutic tools useful in the physiological pathways. This list includes, but is not limited to, the following examples.

**Two-Hybrid System**

[0026] The principles and methods of the yeast two-hybrid system have been described in detail elsewhere (e.g., Bartel and Fields, 1997; Bartel et al., 1993; Fields and Song, 1989; Chevray and Nathans, 1992). The following is a description of the use of this system to identify proteins that interact with a protein of interest.

[0027] The target protein is expressed in yeast as a fusion to the DNA-binding domain of the yeast Gal4p. DNA encoding the target protein or a fragment of this protein is amplified from cDNA by PCR or prepared from an available clone. The resulting DNA fragment is cloned by ligation or recombination into a DNA-binding domain vector (e.g., pGBT9, pGBT.C, pAS2-1) such that an in-frame fusion between the Gal4p and target protein sequences is created.

[0028] The target gene construct is introduced, by transformation, into a haploid yeast strain. A library of activation domain fusions (i.e., adult brain cDNA cloned into an activation domain vector) is introduced by transformation into a haploid yeast strain of the opposite mating type. The yeast strain that carries the activation domain constructs contains one or more Gal4p-responsive reporter gene(s), whose expression can be monitored. Examples of some yeast reporter strains include Y190, PJ69, and CBY14a. An aliquot of yeast carrying the target gene construct is combined with an aliquot of yeast carrying the activation domain library. The two yeast strains mate to form diploid yeast and are plated on media that selects for expression of one or more Gal4p-responsive reporter genes. Colonies that arise after incubation are selected for further characterization.

[0029] The activation domain plasmid is isolated from each colony obtained in the two-hybrid search. The sequence of the insert in this construct is obtained by the dideoxy nucleotide chain termination method. Sequence information is used to identify the gene/protein encoded by the activation domain insert via analysis of the public nucleotide and protein databases. Interaction of the activation domain fusion with the target protein is confirmed by testing for the specificity of the interaction. The activation domain construct is co-transformed into a yeast reporter strain with either the original target protein construct or a variety of other DNA-binding domain constructs. Expression of the reporter genes in the presence of the target protein but not with other test proteins indicates that the interaction is genuine.

[0030] In addition to the yeast two-hybrid system, other genetic methodologies are available for the discovery or detection of protein-protein interactions. For example, a mammalian two-hybrid

system is available commercially (Clontech, Inc.) that operates on the same principle as the yeast two-hybrid system. Instead of transforming a yeast reporter strain, plasmids encoding DNA-binding and activation domain fusions are transfected along with an appropriate reporter gene (e.g., lacZ) into a mammalian tissue culture cell line. Because transcription factors such as the *Saccharomyces cerevisiae* Gal4p are functional in a variety of different eukaryotic cell types, it would be expected that a two-hybrid assay could be performed in virtually any cell line of eukaryotic origin (e.g., insect cells (SF9), fungal cells, worm cells, etc.). Other genetic systems for the detection of protein-protein interactions include the so-called SOS recruitment system (Aronheim et al., 1997).

**Protein-protein interactions**

[0031] Protein interactions are detected in various systems including the yeast two-hybrid system, affinity chromatography, co-immunoprecipitation, subcellular fractionation and isolation of large molecular complexes. Each of these methods is well characterized and can be readily performed by one skilled in the art. See, e.g., U.S. Patents No. 5,622,852 and 5,773,218, and PCT published applications No. WO 97/27296 and WO 99/65939, each of which are incorporated herein by reference.

[0032] The protein of interest can be produced in eukaryotic or prokaryotic systems. A cDNA encoding the desired protein is introduced in an appropriate expression vector and transfected in a host cell (which could be bacteria, yeast cells, insect cells, or mammalian cells). Purification of the expressed protein is achieved by conventional biochemical and immunochemical methods well known to those skilled in the art. The purified protein is then used for affinity chromatography studies: it is immobilized on a matrix and loaded on a column. Extracts from cultured cells or homogenized tissue samples are then loaded on the column in appropriate buffer, and non-binding proteins are eluted. After extensive washing, binding proteins or protein complexes are eluted using various methods such as a gradient of pH or a gradient of salt concentration. Eluted proteins can then be separated by two-dimensional gel electrophoresis, eluted from the gel, and identified by micro-sequencing. The purified proteins can also be used for affinity chromatography to purify interacting proteins disclosed herein. All of these methods are well known to those skilled in the art.

[0033] Similarly, both proteins of the complex of interest (or interacting domains thereof) can be produced in eukaryotic or prokaryotic systems. The proteins (or interacting domains) can be under control of separate promoters or can be produced as a fusion protein. The fusion protein

may include a peptide linker between the proteins (or interacting domains) which, in one embodiment, serves to promote the interaction of the proteins (or interacting domains). All of these methods are also well known to those skilled in the art.

[0034] Purified proteins of interest, individually or a complex, can also be used to generate antibodies in rabbit, mouse, rat, chicken, goat, sheep, pig, guinea pig, bovine, and horse. The methods used for antibody generation and characterization are well known to those skilled in the art. Monoclonal antibodies are also generated by conventional techniques. Single chain antibodies are further produced by conventional techniques.

[0035] DNA molecules encoding proteins of interest can be inserted in the appropriate expression vector and used for transfection of eukaryotic cells such as bacteria, yeast, insect cells, or mammalian cells, following methods well known to those skilled in the art. Transfected cells expressing both proteins of interest are then lysed in appropriate conditions, one of the two proteins is immunoprecipitated using a specific antibody, and analyzed by polyacrylamide gel electrophoresis. The presence of the binding protein (co-immunoprecipitated) is detected by immunoblotting using an antibody directed against the other protein. Co-immunoprecipitation is a method well known to those skilled in the art.

[0036] Transfected eukaryotic cells or biological tissue samples can be homogenized and fractionated in appropriate conditions that will separate the different cellular components. Typically, cell lysates are run on sucrose gradients, or other materials that will separate cellular components based on size and density. Subcellular fractions are analyzed for the presence of proteins of interest with appropriate antibodies, using immunoblotting or immunoprecipitation methods. These methods are all well known to those skilled in the art.

**Disruption of protein-protein interactions**

[0037] It is conceivable that agents that disrupt protein-protein interactions can be beneficial in many physiological disorders, including, but not-limited to NIDDM, AD and others disclosed herein. Each of the methods described above for the detection of a positive protein-protein interaction can also be used to identify drugs that will disrupt said interaction. As an example, cells transfected with DNAs coding for proteins of interest can be treated with various drugs, and co-immunoprecipitations can be performed. Alternatively, a derivative of the yeast two-hybrid system, called the reverse yeast two-hybrid system (Leanna and Hannink, 1996), can be used, provided that the two proteins interact in the straight yeast two-hybrid system.

**Modulation of protein-protein interactions**

[0038] Since the interactions described herein are involved in a physiological pathway, the identification of agents which are capable of modulating the interactions will provide agents which can be used to track physiological disorder or to use lead compounds for development of therapeutic agents. An agent may modulate expression of the genes of interacting proteins, thus affecting interaction of the proteins. Alternatively, the agent may modulate the interaction of the proteins. The agent may modulate the interaction of wild-type with wild-type proteins, wild-type with mutant proteins, or mutant with mutant proteins. Agents which may be used to modulate the protein interaction inlcude a peptide, an antibody, a nucleic acid, an antisense compound or a ribozyme. The nucleic acid may encode the antibody or the antisense compound. The peptide may be at least 4 amino acids of the sequence of either of the interacting proteins. Alternatively, the peptide may be from 4 to 30 amino acids (or from 8 to 20 amino acids) that is at least 75% identical to a contiguous span of amino acids of either of the interacting proteins. The peptide may be covalently linked to a transporter capable of increasing cellular uptake of the peptide. Examples of a suitable transporter include penetratins, *l*-$Tat_{49-57}$, *d*-$Tat_{49-57}$, retro-inverso isomers of *l*- or *d*-$Tat_{49-57}$, L-arginine oligomers, D- arginine oligomers, L-lysine oligomers, D-lysine oligomers, L-histine oligomers, D-histine oligomers, L-ornithine oligomers, D-ornithine oligomers, short peptide sequences derived from fibroblast growth factor, Galparan, and HSV-1 structural protein VP22, and peptoid analogs thereof. Agents can be tested using transfected host cells, cell lines, cell models or animals, such as described herein, by techniques well known to those of ordinary skill in the art, such as disclosed in U.S. Patents Nos. 5,622,852 and 5,773,218, and PCT published application Nos. WO 97/27296 and WO 99/65939, each of which are incorporated herein by reference. The modulating effect of the agent can be tested *in vivo* or *in vitro*. Agents can be provided for testing in a phage display library or a combinatorial library. Exemplary of a method to screen agents is to measure the effect that the agent has on the formation of the protein complex.

**Mutation screening**

[0039] The proteins disclosed in the present invention interact with one or more proteins known to be involved in a physiological pathway, such as in NIDDM, AD or pathways described herein. Mutations in interacting proteins could also be involved in the development of the physiological disorder, such as NIDDM, AD or disorders described herein, for example, through

a modification of protein-protein interaction, or a modification of enzymatic activity, modification of receptor activity, or through an unknown mechanism. Therefore, mutations can be found by sequencing the genes for the proteins of interest in patients having the physiological disorder, such as insulin, and non-affected controls. A mutation in these genes, especially in that portion of the gene involved in protein interactions in the physiological pathway, can be used as a diagnostic tool and the mechanistic understanding the mutation provides can help develop a therapeutic tool.

**Screening for at-risk individuals**

[0040] Individuals can be screened to identify those at risk by screening for mutations in the protein disclosed herein and identified as described above. Alternatively, individuals can be screened by analyzing the ability of the proteins of said individual disclosed herein to form natural complexes. Further, individuals can be screened by analyzing the levels of the complexes or individual proteins of the complexes or the mRNA encoding the protein members of the complexes. Techniques to detect the formation of complexes, including those described above, are known to those skilled in the art. Techniques and methods to detect mutations are well known to those skilled in the art. Techniques to detect the level of the complexes, proteins or mRNA are well known to those skilled in the art.

**Cellular models of Physiological Disorders**

[0041] A number of cellular models of many physiological disorders or diseases have been generated. The presence and the use of these models are familiar to those skilled in the art. As an example, primary cell cultures or established cell lines can be transfected with expression vectors encoding the proteins of interest, either wild-type proteins or mutant proteins. The effect of the proteins disclosed herein on parameters relevant to their particular physiological disorder or disease can be readily measured. Furthermore, these cellular systems can be used to screen drugs that will influence those parameters, and thus be potential therapeutic tools for the particular physiological disorder or disease. Alternatively, instead of transfecting the DNA encoding the protein of interest, the purified protein of interest can be added to the culture medium of the cells under examination, and the relevant parameters measured.

**Animal models**

[0042] The DNA encoding the protein of interest can be used to create animals that overexpress said protein, with wild-type or mutant sequences (such animals are referred to as "transgenic"), or animals which do not express the native gene but express the gene of a second animal (referred to as "transplacement"), or animals that do not express said protein (referred to as "knock-out"). The knock-out animal may be an animal in which the gene is knocked out at a determined time. The generation of transgenic, transplacement and knock-out animals (normal and conditioned) uses methods well known to those skilled in the art.

[0043] In these animals, parameters relevant to the particular physiological disorder can be measured. These parametes may include receptor function, protein secretion *in vivo* or *in vitro*, survival rate of cultured cells, concentration of particular protein in tissue homogenates, signal transduction, behavioral analysis, protein synthesis, cell cycle regulation, transport of compounds across cell or nuclear membranes, enzyme activity, oxidative stress, production of pathological products, and the like. The measurements of biochemical and pathological parameters, and of behavioral parameters, where appropriate, are performed using methods well known to those skilled in the art. These transgenic, transplacement and knock-out animals can also be used to screen drugs that may influence the biochemical, pathological, and behavioral parameters relevant to the particular physiological disorder being studied. Cell lines can also be derived from these animals for use as cellular models of the physiological disorder, or in drug screening.

**Rational drug design**

[0044] The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. Several approaches for use in rational drug design include analysis of three-dimensional structure, alanine scans, molecular modeling and use of anti-id antibodies. These techniques are well known to those skilled in the art. Such techniques may include providing atomic coordinates defining a three-dimensional structure of a protein complex formed by said first polypeptide and said second polypeptide, and designing or selecting compounds capable of interfering with the interaction between a first polypeptide and a second polypeptide based on said atomic coordinates.

[0045] Following identification of a substance which modulates or affects polypeptide activity, the substance may be further investigated. Furthermore, it may be manufactured and/or used in preparation, i.e., manufacture or formulation, or a composition such as a medicament, pharmaceutical composition or drug. These may be administered to individuals.

[0046] A substance identified as a modulator of polypeptide function may be peptide or non-peptide in nature. Non-peptide "small molecules" are often preferred for many *in vivo* pharmaceutical uses. Accordingly, a mimetic or mimic of the substance (particularly if a peptide) may be designed for pharmaceutical use.

[0047] The designing of mimetics to a known pharmaceutically active compound is a known approach to the development of pharmaceuticals based on a "lead" compound. This approach might be desirable where the active compound is difficult or expensive to synthesize or where it is unsuitable for a particular method of administration, e.g., pure peptides are unsuitable active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing is generally used to avoid randomly screening large numbers of molecules for a target property.

[0048] Once the pharmacophore has been found, its structure is modeled according to its physical properties, e.g., stereochemistry, bonding, size and/or charge, using data from a range of sources, e.g., spectroscopic techniques, x-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modeling process.

[0049] A template molecule is then selected, onto which chemical groups that mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted thereon can be conveniently selected so that the mimetic is easy to synthesize, is likely to be pharmacologically acceptable, and does not degrade *in vivo*, while retaining the biological activity of the lead compound. Alternatively, where the mimetic is peptide-based, further stability can be achieved by cyclizing the peptide, increasing its rigidity. The mimetic or mimetics found by this approach can then be screened to see whether they have the target property, or to what extent it is exhibited. Further optimization or modification can then be carried out to arrive at one or more final mimetics for *in vivo* or clinical testing.

**Diagnostic Assays**

[0050] The identification of the interactions disclosed herein enables the development of diagnostic assays and kits, which can be used to determine a predisposition to or the existence of a physiological disorder. In one aspect, one of the proteins of the interaction is used to detect the presence of a "normal" second protein (i.e., normal with respect to its ability to interact with the first protein) in a cell extract or a biological fluid, and further, if desired, to detect the quantitative level of the second protein in the extract or biological fluid. The absence of the "normal" second protein would be indicative of a predisposition or existence of the physiological disorder. In a second aspect, an antibody against the protein complex is used to detect the presence and/or quantitative level of the protein complex. The absence of the protein complex would be indicative of a predisposition or existence of the physiological disorder.

**Nucleic Acids and Proteins**

[0051] A nucleic acid or fragment thereof has substantial identity with another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, more preferably at least about 95% of the nucleotide bases, and more preferably at least about 98% of the nucleotide bases. A protein or fragment thereof has substantial identity with another if, optimally aligned, there is an amino acid sequence identity of at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, more ususally at least about 80% identity, preferably at least about 90% identity, more preferably at least about 95% identity, and most preferably at least about 98% identity.

[0052] Identity means the degree of sequence relatedness between two polypeptide or two polynucleotides sequences as determined by the identity of the match between two strings of such sequences. Identity can be readily calculated. While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (*Computational Molecular Biology*, Lesk, A. M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M

Stockton Press, New York, 1991). Methods commonly employed to determine identity between two sequences include, but are not limited to those disclosed in *Guide to Huge Computers*, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipman, D., *SIAM J Applied Math.* **48**:1073 (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Such methods are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG (Genetics Computer Group, Madison Wis.) program package (Devereux, J., et al., *Nucleic Acids Research* **12**(1).387 (1984)), BLASTP, BLASTN, FASTA (Altschul et al. (1990); Altschul et al. (1997)). The well-known Smith Waterman algorithm may also be used to determine identity.

[0053] Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of $30^{o}$C, typically in excess of $37^{o}$C, and preferably in excess of $45^{o}$C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Asubel, 1992; Wetmur and Davidson, 1968.

[0054] The terms "isolated", "substantially pure", and "substantially homogeneous" are used interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain

purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

[0055] Large amounts of the nucleic acids of the present invention may be produced by (a) replication in a suitable host or transgenic animals or (b) chemical synthesis using techniques well known in the art. Constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art.

[0056] The nucleic acid or protein may also be incorporated on a microarray. The preparation and use of microarrays are well known in the art. Generally, the microarray may contain the entire nucleic acid or protein, or it may contain one or more fragments of the nucleic acid or protein. Suitable nucleic acid fragments may include at least 17 nucleotides, at least 21 nucleotides, at least 30 nucleotides or at least 50 nucleotides of the nucleic acid sequence, particularly the coding sequence. Suitable protein fragments may include at least 4 amino acids, at least 8 amino acids, at least 12 amino acids, at least 15 amino acids, at least 17 amino acids or at least 20 amino acids. Thus, the present invention is also directed to such nucleic acid and protein fragments.

## EXAMPLES

[0057] The present invention is further detailed in the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below are utilized.

EXAMPLE 1

Yeast Two-Hybrid System

[0058] The principles and methods of the yeast two-hybrid systems have been described in detail (Bartel and Fields, 1997). The following is thus a description of the particular procedure that we used, which was applied to all proteins.

[0059] The cDNA encoding the bait protein was generated by PCR from brain cDNA. Gene-specific primers were synthesized with appropriate tails added at their 5' ends to allow recombination into the vector pGBTQ. The tail for the forward primer was 5'-GCAGGAAACAGCTATGACCATACAGTCAGCGGCCGCCACC-3' (SEQ ID NO:1) and the tail for the reverse primer was 5'-ACGGCCAGTCGCGTGGAGTGTTATGTCATGCGGCCGCTA-3' (SEQ ID NO:2). The tailed PCR product was then introduced by recombination into the yeast expression vector pGBTQ, which is a close derivative of pGBTC (Bartel et al., 1996) in which the polylinker site has been modified to include M13 sequencing sites. The new construct was selected directly in the yeast J693 for its ability to drive tryptophane synthesis (genotype of this strain: Mat α, ade2, his3, leu2, trp1, URA3::GAL1-lacZ LYS2::GAL1-HIS3 gal4del gal80del cyhR2). In these yeast cells, the bait is produced as a C-terminal fusion protein with the DNA binding domain of the transcription factor Gal4 (amino acids 1 to 147). A total human brain (37 year-old male Caucasian) cDNA library cloned into the yeast expression vector pACT2 was purchased from Clontech (human brain MATCHMAKER cDNA, cat. # HL4004AH), transformed into the yeast strain J692 (genotype of this strain: Mat a, ade2, his3, leu2, trp1, URA3::GAL1-lacZ LYS2::GAL1-HIS3 gal4del gal80del cyhR2), and selected for the ability to drive leucine synthesis. In these yeast cells, each cDNA is expressed as a fusion protein with the transcription activation domain of the transcription factor Gal4 (amino acids 768 to 881) and a 9 amino acid hemagglutinin epitope tag. J693 cells (Mat α type) expressing the bait were then mated with J692 cells (Mat a type) expressing proteins from the brain library. The resulting diploid yeast cells expressing proteins interacting with the bait protein were selected for the ability to synthesize tryptophan, leucine, histidine, and β-galactosidase. DNA was prepared from each clone, transformed by electroporation into *E. coli* strain KC8 (Clontech KC8 electrocompetent cells, cat. # C2023-1), and the cells were selected on ampicillin-containing plates in the absence of either tryptophane (selection for the bait plasmid) or leucine (selection for the brain library plasmid). DNA for both plasmids was prepared and sequenced by di-deoxynucleotide chain termination method. The identity of the bait cDNA insert was confirmed and the cDNA insert from the brain library plasmid was identified using BLAST program against public

nucleotides and protein databases. Plasmids from the brain library (preys) were then individually transformed into yeast cells together with a plasmid driving the synthesis of lamin fused to the Gal4 DNA binding domain. Clones that gave a positive signal after β-galactosidase assay were considered false-positives and discarded. Plasmids for the remaining clones were transformed into yeast cells together with plasmid for the original bait. Clones that gave a positive signal after β-galactosidase assay were considered true positives.

## EXAMPLE 2

### Identification of LXR-alpha/Utrophin Interaction

[0060] A yeast two-hybrid system as described in Example 1 using amino acids 95-277 of LXR-alpha (GenBank (GB) accession no. U22662) as bait was performed. One clone that was identified by this procedure included amino acids 2443-2650 of utophin (GB accession no. X15488).

## EXAMPLE 3

### Identification of LXR-alpha/PN7771 Interaction

[0061] A yeast two-hybrid system as described in Example 1 using amino acids 95-277 of LXR-alpha (GB accession no. U22662) as bait was performed. One clone that was identified by this procedure included amino acids 1747-2047 of PN7771. The DNA sequence and the predicted protein sequence for PN7771 are set forth in Tables 11 and 12, respectively.

TABLE 11

Nucleotide Sequence of PN7771

cttattttgaaaacatttacatagtgattagttaacccaacagaccaatcctgggaagacagccagagcctgcagcaccttagtaacagaaaaactg
ataattaggagaagagacctgtccaagaccaggaacctggaccaaaattgtgccatgttgctttacttaatgagtggccccagtaaaaactgagct
gtatggcagagctgttcacatttatcttctgtgtccacccagttctgctgaaacccctggcaagatcgtggccctgttgtagcttgtcatgttttgaaca
gctgtctatggaaagaaagcaaacacaacctagagcaacattgatttgttttagaaagctcttttattttcagttctggctgtgttcaacatcttagctta
cgtttttcatgttgtaatgatctgccgtatggacgatcacctctaagttagagagttctgtaatttggcttggattaaagatgcttggttagtgaaagctg
ctgctttttttatagtcaaaggactggttctgagagccttgttgcagatggctgaggtcaccgtcccaagggtgtatgtcgtgtttggcatccattgcat
catggcgaaggcatcttcagatgtgcaggtttcaggctttcatcggaaaatccagcacgttaaaaatgaactttgccacatgttgagcttggaggag
gtggccccagtgctgcagcagacattacttcaggacaacctcttgggcagggtacattttgaccaatttaaagaagcattaatactcatcttgtccag
aactctgtcaaatgaagaacactttcaagaaccagactgctcactagaagctcagcccaaatatgttagaggtgggaagcgttacggacgaaggt
ccttgcccgagttccaagagtccgtggaggagtttcctgaagtgacggtgattgagccactggatgaagaagcgcggccttcacacatcccagc
cggtgactgcagtgagcactggaagacgcaacgcagtgaggagtatgaagcggaaggccagttaaggttttggaacccagatgacttgaatgc
ttcacagagtggatcttcccctccccaagactggatagaagagaaactgcaagaagtttgtgaagatttggggatcacccgtgatggtcacctgaa
ccggaagaagctggtctccatctgtgagcagtatggtttacagaatgtggatggagagatgctcgaggaagtattccataatcttgatcctgacggt
acaatgagtgtagaagatttttttctatggtttgtttaaaaatggaaaatctcttacaccatcagcatctactccatatagacaactaaaaaggcacctttc

catgcagtctttcgatgagagtggacgacgtaccacaacctcatcagcaatgacaagtaccattggctttcgggtcttctcctgcctggatgatggg
atgggccatgcatctgtggagagaatactggacacctggcaggaagagggcattgagaacagccaggagatcctgaaggccttggatttcagc
ctcgatggaaacatcaatttgacagaattaacactggcccttgaaaatgaacttttggttaccaagaacagcattcaccaggcggctctggccagct
ttaaggctgaaatccggcatttgttggaacgagttgatcaggtggtcagagaaaaagagaagctacggtcagatctggacaaggccgagaagct
caagtctttaatggcctcggaggtggatgatcaccatgcggccatagagcggcggaatgagtacaacctcaggaaactggatggagagtacaa
ggagcgaatagcagccttaaaaaatgaactccgaaaagagagagagcagatcctgcagcaggcaggcaagcagcgtttagaacttgaacagg
aaattgaaaaggcaaaaacagaagagaactatatccgggaccgccttgccctctctttaaaggaaaacagtcgtctggaaaatgagcttctagaa
aatgcagagaagttggcagaatatgagaatctgacaaacaaacttcagagaaatttggaaaatgtgttagcagaaaagtttggtgacctcgatcct
agcagtgctgagttcttcctgcaagaagagagactgacacagatgagaaatgaatatgagcggcagtgcagggtactacaagaccaagtagatg
aactccagtctgagctggaagaatatcgtgcacaaggcagagtgctcaggcttccgttgaagaactcaccgtcagaagaagttgaggctaacag
cggtggcattgagcccgaacacgggctcggttctgaagaatgcaatccattgaatatgagcattgaggcagagctggtcattgaacagatgaaag
aacaacatcacagggacatatgttgcctcagactggagctcgaagataaagtgcgccattatgaaaagcagctggacgaaaccgtggtcagctg
caagaaggcacaggagaacatgaagcaaaggcatgagaacgaaacgcgcaccttagaaaaacaaataagtgaccttaaaaatgaaattgctga
acttcaggggcaagcagcagtgctcaaggaggcacatcatgaggccacttgcaggcatgaggaggagaaaaaacaactgcaagtgaagcttg
aggaggaaaagactcacctgcaggagaagctgaggctgcaacatgagatggagctcaaggctagactgacacaggctcaagcaagctttgag
cgggagagggaaggccttcagagtagcgcctggacagaagagaaggtgagaggcttgactcaggaactagagcagtttcaccaggagcagc
tgacaagcctggtggagaaacacactcttgagaaagaggagttaagaaaagagctcttggaaaagcaccaaagggagcttcaggagggaagg
gaaaaaatggaaacagagtgtaatagaagaacctctcaaatagaagcccagtttcagtctgattgtcagaaagtcactgagaggtgtgaaagcgc
tctgcaaagcctggaggggcgctaccgccaagagctgaaggacctccaggaacagcagcgtgaggagaaatcccagtgggaatttgagaag
gacgagctcacccaggagtgtgcggaagcccaggagctgctgaaagagactcttaagagagagaaaacaacttctctggtcctgacccagga
gagagagatgctggagaaaacatacaaagaacatttgaacagcatggtcgtcgagagacagcagctactccaagacctggaagacctaagaaa
tgtatctgaaacccagcaaagcctgctgtctgaccagatacttgagctgaagagcagtcacaaaagggaactgagggagcgtgaggaggtcct
gtgccaggcaggggcttcggagcagctggccagccagcggctggaaagactagaaatggaacatgaccaggaaaggcaggaaatgatgtcc
aagcttctagccatggagaacattcacaaagcgacctgtgagacagcagatcgagaaagagccgagatgagcacagaaatctccagacttcag
agtaaaataaaggaaatgcagcaggcaacatctcctctctcaatgcttcagagtggttgccaggtgataggagaggaggaggtggaaggagatg
gagccctgtccctgcttcagcaaggggagcagctgttggaagaaaatggggacgtcctcttaagcctgcagagagctcatgaacaggcagtga
aggaaaatgtgaaaatggctactgaaatttctagattgcaacagaggctacaaaagttagagccagggttagtaatgtcttcttgtttggatgagcca
gctactgagttttttggaaatactgcggaacaaacagagcagtttttacagcaaaaccgaacgaagcaagtagaaggtgtgaccaggcggcatgt
cctaagtgacctggaagatgatgaggtccgggacctgggaagtacagggacgagctctgttcagagacaggaagtcaaaatagaggagtctga
agcttcagtagagggtttttctgagcttgaaaacagtgaagagaccaggactgaatcctgggagctgaagaatcagattagtcagcttcaggaaca
gctaatgatgttatgtgcggactgtgatcgagcttctgaaaagaaacaggacctacttttgatgtttctgtgctaaaaaagaaactgaagatgcttga
gagaatccctgaggcttctcccaaatataagctgttgtatgaagatgtgagccgagaaaatgactgccttcaggaagagctgagaatgatggaga
cacgctacgatgaggcactagaaaataacaaagaactcactgcagaggttttcaggttgcaggatgagctgaagaaaatggaggaagtcactga
aacattcctcagcctggaaaagagttacgatgaggtcaaaatagaaaatgaggggctgaatgttctggttttgagacttcaaggcaagattgagaa
gcttcaggaaagcgtggtccagcggtgtgactgctgcttatgggaagccagtttagagaacctggaaatcgaacctgatggaaatatactccagc
tcaatcagacactggaagagtgtgtgcccagggttaggagtgtacatcatgtcatagaggaatgtaagcaagaaaaccagtaccttgaggggaa
cacacagctcttggaaaaagtaaaagcacatgaaattgcctggttacatggaacaattcagacacatcaagaaaggccaagagtacagaatcaa
gttatactggaggaaaacactactctcctaggctttcaagacaaacattttcagcatcaggccaccatagcagagttagaactggagaaaacaaag
ttacaggagctgactaggaagttgaaggagagagtcactattttagttaagcaaaaagatgtactttctcacggagaaaaggaggaagagctgaa
ggcaatgatgcatgacttgcagatcacgtgcagtgagatgcagcaaaaagttgaacttctgagatatgaatctgaaaagcttcaacaggaaaattc
tattttgagaaatgaaattactactttaaatgaagaagatagcatttctaacctgaaattagggacattaaatggatctcaggaagaaatgtggcaaaa
aacggaaactgtaaaacaagaaaatgctgcagttcagaagatggttgaaaatttaaagaaacagatttcagaattaaaaatcaaaaaccaacaatt
ggatttggaaaatacagaacttagccaaaagaactctcaaaaccaggaaaaactgcaagaacttaatcaacgtctaacagaaatgctatgccaga
aggaaaaagagccaggaaacagtgcattggaggaacgggaacaagagaagtttaatctgaaagaagaactggaacgttgtaaagtgcagtcct
ccactttagtgtcttctctggaggcggagctctctgaagttaaaatacagacccatattgtgcaacaggaaaaccaccttctcaaagatgaactgga
gaaaatgaaacagctgcacagatgtcccgatctctctgacttccagcaaaaaatctctagtgttctaagctacaacgaaaaactgctgaaagaaaa
ggaagctctgagtgaggaattaaatagctgtgtcgataagttggcaaaatcaagtcttttagagcatagaattgcgacgatgaagcaggaacagaa

atcctgggaacatcagagtgcgagcttaaagtcacagctggtggcttctcaggaaaaggttcagaatttagaagacaccgtgcagaatgtaaacc
tgcaaatgtcccggatgaaatctgacctacgagtgactcagcaggaaaaggaggctttaaaacaagaagtgatgtctttacataagcaacttcaga
atgctggtggcaagagctgggccccagagatagctactcatccatcagggctccataaccagcagaaaaggctgtcctgggacaagttggatca
tctgatgaatgaggaacagcagctgctttggcaagagaatgagaggctccagaccatggtacagaacaccaaagccgaactcacgcactcccg
ggagaaggtccgtcaattggaatccaatcttcttcccaagcaccaaaaacatctaaacccatcaggtaccatgaatcccacagagcaagaaaaatt
gagcttaaagagagagtgtgatcagtttcagaaagaacaatctcctgctaacaggaaggtcagtcagatgaattcccttgaacaagaattagaaac
aattcatttggaaaatgaaggcctgaaaaagaaacaagtaaaactggatgagcagctcatggagatgcagcacctgaggtccactgcgacgcct
agcccgtcccctcatgcttgggatttgcagctgctccagcagcaagcctgtccgatggtgcccagggagcagtttctgcagcttcaacgccagct
gctgcaggcagaaaggataaaccagcacctgcaggaggaacttgaaaacaggacctccgaaaccaacacaccacagggaaaccaggaaca
actggtaactgtcatggaggaacgaatgatagaagttgaacagaaactgaaactagtgaaaaggcttcttcaagagaaagtgaatcagctcaaag
aacaactctgcaagaacactaaggcagacgcaatggtgaaggacttgtatgttgaaaatgcccagttgttgaaagctctggaagtgactgaacag
cgacagaaaacagcagagaagaaaaattacctcctggaggagaagattgccagcctcagtaatatagttaggaatctgacaccagcgccattga
cttctacacctcctttgaggtcatagccaaaccaaagggtacactcatatttgtgcactttactgaaatagatgaacatttcagtaggttctcaacttaa
aattaagcctaacctaaaactgccagcaacacaactggagtttccatttatcataattagtttttctaaatagacccttatgggagtttgaaaataaatac
tcacatatttcactacttaaattattcccaagatttgaatttatttttaaaatttttaatagccaccaagaatgtggacatatgaaaattcaagaacctaaaaa
ataccagttttgaatgagtttttgtggttttggtttttttaattattacaaatctatgtgtaaaatctagatatttgaagtttgagatctgatgagaatggttgtta
taaactttatttttaaaaccaaatttaggtgttcttacatatttaaatactggaaagtcattataatagttttggttctttgaattggtagacaattagtagagta
taattggttaggaggcagggcttattaagtggttattaaccgctgacatcagacaaacccaaatctgtagaattctaacctcctaacacctgtgacag
tattaccactcttcttgtattatagatttagaactgatttactcaattgcactcttaactaatgttaaaagcttacttgctttaaacagccttttcttctttctctt
aaaagtttcatttggggagctggtcttctaagaaacggataaagccacataattaaagcagttgaactagagggaaagcactgaacaaaccacttt
ggagtaaatagctactcttagaaaagagggataagcagaccatgtaggttttctgtctctcaaatcttagagttcataaatttacttgaggttgcctcaa
gaactcagggaacaatactgtaaactgtcttcctgaactactgtagggcctctctaagaatttgaaatgtataaaccatgtgacctcatttatttgtctta
tatatttacagccatactagaatttttatttctacgtttttagtaaatttaatattctgggggaaaaaaggccttgattttagggttaaaaacctgacttatag
aagagtttatttaatataggtcaaaatttttctgtgtttcttattccttctatacctcaaatctgattctaagaatttcttactgtgataatcattggcatgccac
ctgaggtcaaggagtgccaaataggactttccactcatgctcaagatcaaaactttatagaacagtcaacattttagattcggtaacctttttttttcttcc
aattataatctctgcttctagccacttccgccagcagttggtggaagacttactaggtgcagggcactttccaagttcatcacaacaacctgcttgtttt
catgagacaataatccgaaaagttcgctttgatatattcctggagggccaagcccatctatttacaaaaggtgaacagcaaaatcaagcactgcttt
atgggcaggaacacaagagaaagcaaactgcccaagaagtcatcatgtcagaaactcaatctcaacaaaataatttccatcagggaacttcagg
gtttcttgggggcttatgagtctcaccggtcaacccaggaggcctcactacaagagccttgacaaggcactgttttttgtgggactgggagttcaca
ctgatgaagcaaacctttgaatttttgcacagctcttgtcagaaagccctgagttccccctggataaagagttaatttttaatccttccctataattatactt
caaaatatttgacatctgctattatgccttctttagatctttcttctgcggtgcagacatttctagtaagtgtttgactacttgtatggcattagctttcacag
aaaattgtttcacttaaaactgtggattggcctaggctaaggacaaaaataaactaagtacctgtagtgtatttatgtgatatgtgtcaagttactcaaa
gttattgctgttggaactgaacaataatatttcccagatagctggccttagcatgtgatcacggttgttgtatttttaatttttgtcttttacagtatgagag
gtgtaggttaatttgtttatttcctataaatttgtatttatgtgtatataaaatgtacaatgaatgtaaatatgactttctggaaagtttagactacatttagaat
ctctattcaaaatcaaaatgctgctcaaatgaatttaaccaacatctaggtgcttaatttctcattttatcccacttatgagattgggaaaaagatcaatat
gagaaataccatacagatacttaaatgtatgcatttgtgcaacaatttttgagaaggtgagtggcaatttataatttagttggcaatttataatagaact
tatagcttttaaaagacttttaaagacattaaatgtaaacttaaaaatgtttagatcttgtttcaaactttacaatagcattcttcaaaatattaagttatatat
tttataggcatttagttgcttattaaaagcactgattttcaaacttttgatttaagaacaattatttaagatcgtctcagaagatgggatcttcgtttcaaga
aaagggaatcaagtttgcctttgagataatacgttacactaagaaaaggaaaatgtggatagtaaaacccacctctctcatcctattgtactctcttct
gctttttagaagcctgcacttaagcttagatttgtgaagggagagtagaaggggagaagtagaaccacagtgttttatttatttttctaaaactcttact
aaatccagatttttaaactgttttaaatgtgaattcttcccagaaatttcaatgcattgcatatttagccttcggcatatttttcatgaatagatcatgaagt
cataggcttccaaggcataggaagagatcttgcaggtctagtatttaataatgcactattacccagggcagatattatgagaaactgtttcttctctaa
gggtttatggcagactttgctttttaacatgtgagaaatgaatttttattttgtgatttatgtgatttcttttgctgagtgaaggaaaggagaaattgttgc
tattgtcagcatcttaaaggtatttccagtcaaggcaaggctaagtgctttgtgatagtattaagcaagtcatgttttgaatggattacctgtagtgactc
attggaatgatataattatacaagtaatgccaaaaaccaagtcaaagcctaattaaccaaagcactcatttaaaaatcatcatgtttggacctatctgg
acctctcagcactgtaaaatagttttggttttgtggcatatgaatagctgtttaacaaatcaaagttagctttttgcttctcagctttttgggcaatacaag
ttaagttcttaatggggagacattatcatggcatgacttaagggaacattggtttgtgaaggaaaaacagattatctaaagccatctctatgtttctgttc

agataaagattaatgagttctgtgtttatatcagctttgtatatttcatcttagccattctatcctagaaagattttaatgtgagcttaagatgtaaataaata attttgcaaacatgaaaaaaaaaaaaaaaa (SEQ ID NO:3)

TABLE 12

Predicted Amino Acid Sequence of PN7771

MAEVTVPRVYVVFGIHCIMAKASSDVQVSGFHRKIQHVKNELCHMLSLEEVAPVLQQTLL
QDNLLGRVHFDQFKEALILILSRTLSNEEHFQEPDCSLEAQPKYVRGGKRYGRRSLPEFQES
VEEFPEVTVIEPLDEEARPSHIPAGDCSEHWKTQRSEEYEAEGQLRFWNPDDLNASQSGSSP
PQDWIEEKLQEVCEDLGITRDGHLNRKKLVSICEQYGLQNVDGEMLEEVFHNLDPDGTMS
VEDFFYGLFKNGKSLTPSASTPYRQLKRHLSMQSFDESGRRTTTSSAMTSTIGFRVFSCLDD
GMGHASVERILDTWQEEGIENSQEILKALDFSLDGNINLTELTLALENELLVTKNSIHQAAL
ASFKAEIRHLLERVDQVVREKEKLRSDLDKAEKLKSLMASEVDDHHAAIERRNEYNLRKL
DGEYKERIAALKNELRKEREQILQQAGKQRLELEQEIEKAKTEENYIRDRLALSLKENSRLE
NELLENAEKLAEYENLTNKLQRNLENVLAEKFGDLDPSSAEFFLQEERLTQMRNEYERQCR
VLQDQVDELQSELEEYRAQGRVLRLPLKNSPSEEVEANSGGIEPEHGLGSEECNPLNMSIEA
ELVIEQMKEQHHRDICCLRLELEDKVRHYEKQLDETVVSCKKAQENMKQRHENETRTLEK
QISDLKNEIAELQGQAAVLKEAHHEATCRHEEEKKQLQVKLEEEKTHLQEKLRLQHEMEL
KARLTQAQASFEREREGLQSSAWTEEKVRGLTQELEQFHQEQLTSLVEKHTLEKEELRKEL
LEKHQRELQEGREKMETECNRRTSQIEAQFQSDCQKVTERCESALQSLEGRYRQELKDLQE
QQREEKSQWEFEKDELTQECAEAQELLKETLKREKTTSLVLTQEREMLEKTYKEHLNSMV
VERQQLLQDLEDLRNVSETQQSLLSDQILELKSSHKRELREREEVLCQAGASEQLASQRLER
LEMEHDQERQEMMSKLLAMENIHKATCETADRERAEMSTEISRLQSKIKEMQQATSPLSM
LQSGCQVIGEEEVEGDGALSLLQQGEQLLEENGDVLLSLQRAHEQAVKENVKMATEISRLQ
QRLQKLEPGLVMSSCLDEPATEFFGNTAEQTEQFLQQNRTKQVEGVTRRHVLSDLEDDEVR
DLGSTGTSSVQRQEVKIEESEASVEGFSELENSEETRTESWELKNQISQLQEQLMMLCADCD
RASEKKQDLLFDVSVLKKKLKMLERIPEASPKYKLLYEDVSRENDCLQEELRMMETRYDE
ALENNKELTAEVFRLQDELKKMEEVTETFLSLEKSYDEVKIENEGLNVLVLRLQGKIEKLQE
SVVQRCDCCLWEASLENLEIEPDGNILQLNQTLEECVPRVRSVHHVIEECKQENQYLEGNT
QLLEKVKAHEIAWLHGTIQTHQERPRVQNQVILEENTTLLGFQDKHFQHQATIAELELEKTK
LQELTRKLKERVTILVKQKDVLSHGEKEEELKAMMHDLQITCSEMQQKVELLRYESEKLQ
QENSILRNEITTLNEEDSISNLKLGTLNGSQEEMWQKTETVKQENAAVQKMVENLKKQISE
LKIKNQQLDLENTELSQKNSQNQEKLQELNQRLTEMLCQKEKEPGNSALEEREQEKFNLKE
ELERCKVQSSTLVSSLEAELSEVKIQTHIVQQENHLLKDELEKMKQLHRCPDLSDFQQKISS
VLSYNEKLLKEKEALSEELNSCVDKLAKSSLLEHRIATMKQEQKSWEHQSASLKSQLVASQ
EKVQNLEDTVQNVNLQMSRMKSDLRVTQQEKEALKQEVMSLHKQLQNAGGKSWAPEIAT
HPSGLHNQQKRLSWDKLDHLMNEEQQLLWQENERLQTMVQNTKAELTHSREKVRQLESN
LLPKHQKHLNPSGTMNPTEQEKLSLKRECDQFQKEQSPANRKVSQMNSLEQELETIHLENE
GLKKKQVKLDEQLMEMQHLRSTATPSPSPHAWDLQLLQQQACPMVPREQFLQLQRQLLQ
AERINQHLQEELENRTSETNTPQGNQEQLVTVMEERMIEVEQKLKLVKRLLQEKVNQLKEQ
LCKNTKADAMVKDLYVENAQLLKALEVTEQRQKTAEKKNYLLEEKIASLSNIVRNLTPAPL
TSTPPLRS (SEQ ID NO:4)

EXAMPLES 4-12

Identification of Protein-Protein Interactions

[0062] A yeast two-hybrid system as described in Example 1 using amino acids of the bait as set forth in Table 13 was performed. The clone that was identified by this procedure for each bait is set forth in Table 13 as the prey. The "AA" refers to the amino acids of the bait or prey. The "NUC" refers to the nucleotides of the bait or prey. The Accession numbers refer to GB: GenBank accession numbers.

TABLE 13

| Ex. | BAIT | ACCESSION | COORDINATES | PREY | ACCESSION | COORDINATES |
|---|---|---|---|---|---|---|
| 4 | LXR-alpha | GB: U22662 | AA: 95-156 | zyxin | GB: X94991 | AA 323-572 |
| 5 | LXR-alpha | GB: U22662 | AA 95-156 | LIMS1 | GB: U09284 | AA 25-90 |
| 6 | LXR-alpha | GB: U22662 | AA 95-277 | Homer-3 | GB: NM_004838 | AA 182-354 |
| 7 | LXR-alpha | GB: U22662 | AA 95-277 | RACK1 | GB: M24194 | AA 178-317 |
| 8 | LXR-alpha | GB: U22662 | AA 257-448 | EIF3S1 | GB: NM_003758 | AA 94-254 |
| 9 | LXR-alpha | GB: U22662 | AA 156-448 | EIF3S1 | GB: NM_003758 | AA 94-254 |
| 10 | LXR-alpha | GB: U22662 | AA 95-277 | PSMD11 | GB: NM_002815 | AA 87-422 |
| 11 | LXR-alpha | GB: U22662 | AA 257-448 | KIAA0610 | GB: AB011182 | AA 36-245 |
| 12 | LXR-alpha | GB: U22662 | AA 156-448 | CIR | GB: U03644 | AA 226-450 |

EXAMPLE 13

Generation of Polyclonal Antibody Against Protein Complexes

[0063] As shown above, LXR-alpha interacts with utrophin to form a complex. A complex of the two proteins is prepared, e.g., by mixing purified preparations of each of the two proteins. If desired, the protein complex can be stabilized by cross-linking the proteins in the complex, by methods known to those of skill in the art. The protein complex is used to immunize rabbits and mice using a procedure similar to that described by Harlow et al. (1988). This procedure has been shown to generate Abs against various other proteins (for example, see Kraemer et al., 1993).

[0064] Briefly, purified protein complex is used as immunogen in rabbits. Rabbits are immunized with 100 μg of the protein in complete Freund's adjuvant and boosted twice in three-week intervals, first with 100 μg of immunogen in incomplete Freund's adjuvant, and followed by 100 μg of immunogen in PBS. Antibody-containing serum is collected two weeks thereafter. The antisera is preadsorbed with LXR-alpha and utrophin, such that the remaining antisera comprises antibodies which bind conformational epitopes, i.e., complex-specific epitopes, present on the LXR-alpha-utrophin complex but not on the monomers.

[0065] Polyclonal antibodies against each of the complexes set forth in Tables 1-10 are prepared in a similar manner by mixing the specified proteins together, immunizing an animal and isolating antibodies specific for the protein complex, but not for the individual proteins.

[0066] Polyclonal antibodies against the protein set forth in Table 12 are prepared in a similar manner by immunizing an animal with the protein and isolating antibodies specific for the protein.

EXAMPLE 14

Generation of Monoclonal Antibodies Specific for Protein Complexes

[0067] Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising LXR-alpha/utrophin complexes conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known in the art. The complexes can be prepared as described in Example 13, and may also be stabilized by cross-linking. The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 μg of immunogen, and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice

with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

[0068] Spleens are removed from immune mice and a single-cell suspension is prepared (Harlow et al., 1988). Cell fusions are performed essentially as described by Kohler et al. (1975). Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, MD) or NS-1 myeloma cells are fused with immune spleen cells using polyethylene glycol as described by Harlow et al. (1988). Cells are plated at a density of $2x10^5$ cells/well in 96-well tissue culture plates. Individual wells are examined for growth, and the supernatants of wells with growth are tested for the presence of LXR-alpha/utrophin complex-specific antibodies by ELISA or RIA using LXR-alpha/utrophin complex as target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

[0069] Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibodies for characterization and assay development. Antibodies are tested for binding to LXR-alpha alone or to utrophin alone, to determine which are specific for the LXR-alpha/utrophin complex as opposed to those that bind to the individual proteins.

[0070] Monoclonal antibodies against each of the complexes set forth in Tables 1-10 are prepared in a similar manner by mixing the specified proteins together, immunizing an animal, fusing spleen cells with myeloma cells and isolating clones which produce antibodies specific for the protein complex, but not for the individual proteins.

[0071] Monoclonal antibodies against the protein set forth in Table 12 are prepared in a similar manner by immunizing an animal with the protein, fusing spleen cells with myeloma cells and isolating clones which produce antibodies specific for the protein.

## EXAMPLE 15

### *In vitro* Identification of Modulators for Protein-Protein Interactions

[0072] The present invention is useful in screening for agents that modulate the interaction of LXR-alpha and utrophin. The knowledge that LXR-alpha and utrophin form a complex is useful in designing such assays. Candidate agents are screened by mixing LXR-alpha and utrophin (a) in the presence of a candidate agent, and (b) in the absence of the candidate agent. The amount of complex formed is measured for each sample. An agent modulates the interaction of LXR-alpha and utrophin if the amount of complex formed in the presence of the agent is greater than

(promoting the interaction), or less than (inhibiting the interaction) the amount of complex formed in the absence of the agent. The amount of complex is measured by a binding assay, which shows the formation of the complex, or by using antibodies immunoreactive to the complex.

[0073] Briefly, a binding assay is performed in which immobilized LXR-alpha is used to bind labeled utrophin. The labeled utrophin is contacted with the immobilized LXR-alpha under aqueous conditions that permit specific binding of the two proteins to form a LXR-alpha/utrophin complex in the absence of an added test agent. Particular aqueous conditions may be selected according to conventional methods. Any reaction condition can be used as long as specific binding of LXR-alpha/utrophin occurs in the control reaction. A parallel binding assay is performed in which the test agent is added to the reaction mixture. The amount of labeled utrophin bound to the immobilized LXR-alpha is determined for the reactions in the absence or presence of the test agent. If the amount of bound, labeled utrophin in the presence of the test agent is different than the amount of bound labeled utrophin in the absence of the test agent, the test agent is a modulator of the interaction of LXR-alpha and utrophin.

[0074] Candidate agents for modulating the interaction of each of the protein complexes set forth in Tables 1-10 are screened *in vitro* in a similar manner.

## EXAMPLE 16

### *In vivo* Identification of Modulators for Protein-Protein Interactions

[0075] In addition to the *in vitro* method described in Example 15, an *in vivo* assay can also be used to screen for agents which modulate the interaction of LXR-alpha and utrophin. Briefly, a yeast two-hybrid system is used in which the yeast cells express (1) a first fusion protein comprising LXR-alpha or a fragment thereof and a first transcriptional regulatory protein sequence, e.g., GAL4 activation domain, (2) a second fusion protein comprising utrophin or a fragment thereof and a second transcriptional regulatory protein sequence, e.g., GAL4 DNA-binding domain, and (3) a reporter gene, e.g., β-galactosidase, which is transcribed when an intermolecular complex comprising the first fusion protein and the second fusion protein is formed. Parallel reactions are performed in the absence of a test agent as the control and in the presence of the test agent. A functional LXR-alpha/utrophin complex is detected by detecting the amount of reporter gene expressed. If the amount of reporter gene expression in the presence of the test agent is different than the amount of reporter gene expression in the absence of the test agent, the test agent is a modulator of the interaction of LXR-alpha and utrophin.

[0076] Candidate agents for modulating the interaction of each of the protein complexes set forth in Tables 1-10 are screened *in vivo* in a similar manner.

[0077] While the invention has been disclosed in this patent application by reference to the details of preferred embodiments of the invention, it is to be understood that the disclosure is intended in an illustrative rather than in a limiting sense, as it is contemplated that modifications will readily occur to those skilled in the art, within the spirit of the invention and the scope of the appended claims.

## BIBLIOGRAPHY

Aronheim et al., 1997. Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Mol. Cell. Biol.* **17**:3094-3102.

Altschul, S.F. *et al.* (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

Altschul, S.F. *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**:3389-3402.

Banres, P.J. (1998). Anti-inflammatory actions of glucocorticoids: molecular mechanisms. *Clinical Science* **94**:557-572.

Bartel, P.L. et al. (1993). "Using the 2-hybrid system to detect protein-protein interactions." In: *Cellular Interactions in Development: A Practical Approach,* Oxford University Press, pp. 153-179.

Bartel, P.L. et al. (1996). A protein linkage map of Escherichia coli bacteriophage T7. *Nat Genet* **12**:72-77.

Bartel, P.L. and Fields, S. (1997). *The Yeast Two-Hybrid System.* New York: Oxford University Press.

Brown, A.J. and Jessup, W. (1999). Oxysterols and atherosclerosis. *Atherosclerosis* **142**:1-28.

Chevray, P.M. and Nathans, D.N. (1992). Protein interaction cloning in yeast: identification of mammalian proteins that interact with the leucine zipper of Jun. *Proc. Natl. Acad. Sci. USA* **89**:5789-5793.

DeFranco, D. et al. (1998). Molecular chaperones and subcellular trafficking of steroid receptors. *J. Steroid Biochem. and Mol. Biol.* **65**:51-58.

Duprez, D. et al. (2000). Aldosterone and vascular damage. *Curr. Hypertens. Rep.* **2**:327-34.

Fields S and Song O-K (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**:245-246.

Geller, D.S. et al. (2000). Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* **289**:119-123.

Harlow et al. (1988). *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Haynes, M.P. et al. (2000). Molecular mechanisms of estrogen actions ion the vasculature. *J. Nucl. Cardiol.* 7:500-508.

Hong, Y.R. et al. (2000). Cloning and characterization of a novel human ninein protein that interacts with the glycogen synthase kinase 3beta. *Biochim Biophys Acta* **1492**:513-6.

Hsieh, J.J. et al. (1999). CIR, a corepressor linking the DNA binding factor CBF1 to the histone deacetylase complex. *Proc Natl Acad Sci USA.* **96**:23-8.

Klein-Hitpass, L. et al. (1998). Targets of activated steroid hormone receptors: basal transcription factors and receptor interacting proteins. *J. Mol. Med.* **76**:490-496.

Kohler, G. and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**:495-497.

Kraemer, F.B. et al. (1993). Detection of hormone-sensitive lipase in various tissues. I. Expression of an HSL/bacterial fusion protein and generation of anti-HSL antibodies. *J. Lipid Res.* **34**:663-672.

Leanna, C.A. and Hannink, M. (1996). The reverse two-hybrid system: a genetic scheme for selection against specific protein/protein interactions. *Nucl. Acids Res.* **24**:3341-3347.

Mack, G.J. et al. (1998). Autoantibodies to a group of centrosomal proteins in human autoimmune sera reactive with the centrosome. *Arthritis Rheum.* **41**:551-8.

Mangelsdorf, D. et al. (1995). The nuclear receptor superfamily: the second decade. *Cell* **83**:835-839.

Moras, D. and Gronemeyer, H. (1998). The nuclear receptor ligand-binding domain: structure and function. *Curr. Op. Cell Biol.* **10**:384-391.

Schroepfer, G.J. (2000). Oxysterols: modulators of cholesterol metabolism and other processes. *Physiol. Rev.* **80**:361-554.

Tu, Y. et al. (1998). Nck-2, a novel Src homology2/3-containing adaptor protein that interacts with the LIM-only protein PINCH and components of growth factor receptor kinase-signaling pathways. *Mol Biol Cell.* **9**:3367-82.

Wetmur, J.G. and Davidson, N. (1968). "Kinetics of renaturation of DNA." *J. Mol. Biol.* **31**:349-370.

Whitfield, G. et al. (1999). Steroid hormone receptors: evolution, ligands, and molecular basis of biologic function. *J. Biol. Chem.* **Suppl 32-33**:110-122.

Wolffe, A. et al. (1997). Chromatin remodeling by steroid and nuclear receptors. *Cell Research* 7:127-142.

PCT Published Application No. WO 97/27296

PCT Published Application No. WO 99/65939

U.S. Patent No. 5,622,852

U.S. Patent No. 5,773,218

WHAT IS CLAIMED IS:

1. An isolated protein complex comprising two proteins, the protein complex selected from the group consisting of:

(i) a complex of a first protein and a second protein;

(ii) a complex of a fragment of said first protein and said second protein;

(iii) a complex of said first protein and a fragment of said second protein; and

(iv) a complex of a fragment of said first protein and a fragment of said second protein, wherein said first protein is LXR-alpha and said second protein is selected from the group consisting of utrophin, zyxin, LIMS1, PN7771, Homer-3, RACK1, EIF3S1, PSMD11, KIAA0610 and CIR.

2. The protein complex of claim 1, wherein said protein complex comprises said first protein and said second protein.

3. The protein complex of claim 1, wherein said protein complex comprises a fragment of said first protein and said second protein or said first protein and a fragment of said second protein.

4. The protein complex of claim 1, wherein said protein complex comprises fragments of said first protein and said second protein.

5. An isolated antibody selectively immunoreactive with a protein complex of claim 1.

6. The antibody of claim 5, wherein said antibody is a monoclonal antibody.

7. A method for diagnosing a physiological disorder in an animal, which comprises assaying for:

(a) whether a protein complex set forth in claim 1 is present in a tissue extract;

(b) the ability of proteins to form a protein complex set forth in claim 1; and

(c) a mutation in a gene encoding a protein of a protein complex set forth in claim 1.

8. The method of claim 7, wherein said animal is a human.

9. The method of claim 8, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

10. The method of claim 7, wherein the diagnosis is for a predisposition to said physiological disorder.

11. The method of claim 7, wherein the diagnosis is for the existence of said physiological disorder.

12. The method of claim 7, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

13. The method of claim 7, wherein said assay comprises a yeast two-hybrid assay.

14. The method of claim 7, wherein said assay comprises measuring *in vitro* a complex formed by combining the proteins of the protein complex, said proteins isolated from said animal.

15. The method of claim 14, wherein said complex is measured by binding with an antibody specific for said complex.

16. The method of claim 7, wherein said assay comprises mixing an antibody specific for said protein complex with a tissue extract from said animal and measuring the binding of said antibody.

17. A method for determining whether a mutation in a gene encoding one of the proteins of a protein complex set forth in claim 1 is useful for diagnosing a physiological disorder, which comprises assaying for the ability of said protein with said mutation to form a complex with the other protein of said protein complex, wherein an inability to form said complex is indicative of said mutation being useful for diagnosing a physiological disorder.

18. The method of claim 17, wherein said gene is an animal gene.

19. The method of claim 18, wherein said animal is a human.

20. The method of claim 19, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

21. The method of claim 17, wherein the diagnosis is for a predisposition to a physiological disorder.

22. The method of claim 17, wherein the diagnosis is for the existence of a physiological disorder.

23. The method of claim 17, wherein said assay comprises a yeast two-hybrid assay.

24. The method of claim 17, wherein said assay comprises measuring *in vitro* a complex formed by combining the proteins of the protein complex, said proteins isolated from an animal.

25. The method of claim 24, wherein said animal is a human.

26. The method of claim 24, wherein said complex is measured by binding with an antibody specific for said complex.

27. A non-human animal model for a physiological disorder wherein the genome of said animal or an ancestor thereof has been modified such that the formation of a protein complex set forth in claim 1 has been altered.

28. The non-human animal model of claim 27, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

29. The non-human animal model of claim 27, wherein the formation of said protein complex has been altered as a result of:

(a) over-expression of at least one of the proteins of said protein complex;

(b) replacement of a gene for at least one of the proteins of said protein complex with a gene from a second animal and expression of said protein;

(c) expression of a mutant form of at least one of the proteins of said protein complex;

(d) a lack of expression of at least one of the proteins of said protein complex; or

(e) reduced expression of at least one of the proteins of said protein complex.

30. A cell line obtained from the animal model of claim 27.

31. A non-human animal model for a physiological disorder, wherein the biological activity of a protein complex set forth in claim 1 has been altered.

32. The non-human animal model of claim 31, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

33. The non-human animal model of claim 31, wherein said biological activity has been altered as a result of:

(a) disrupting the formation of said complex; or

(b) disrupting the action of said complex.

34. The non-human animal model of claim 31, wherein the formation of said complex is disrupted by binding an antibody to at least one of the proteins which form said protein complex.

35. The non-human animal model of claim 31, wherein the action of said complex is disrupted by binding an antibody to said complex.

36. The non-human animal model of claim 31, wherein the formation of said complex is disrupted by binding a small molecule to at least one of the proteins which form said protein complex.

37. The non-human animal model of claim 31, wherein the action of said complex is disrupted by binding a small molecule to said complex.

38. A cell in which the genome of cells of said cell line has been modified to produce at least one protein complex set forth in claim 1.

39. A cell line in which the genome of the cells of said cell line has been modified to eliminate at least one protein of a protein complex set forth in claim 1.

40. A composition comprising:

a first expression vector having a nucleic acid encoding a first protein or a homologue or derivative or fragment thereof; and

a second expression vector having a nucleic acid encoding a second protein, or a homologue or derivative or fragment thereof, wherein said first and said second proteins are the proteins of claim 1.

41. A host cell comprising:

a first expression vector having a nucleic acid encoding a first protein which is first protein or a homologue or derivative or fragment thereof; and

a second expression vector having a nucleic acid encoding a second protein which is second protein, or a homologue or derivative or fragment thereof thereof, wherein said first and said second proteins are the proteins of claim 1.

42. The host cell of claim 41, wherein said host cell is a yeast cell.

43. The host cell of claim 41, wherein said first and second proteins are expressed in fusion proteins.

44. The host cell of claim 41, wherein one of said first and second nucleic acids is linked to a nucleic acid encoding a DNA binding domain, and the other of said first and second nucleic acids is linked to a nucleic acid encoding a transcription-activation domain, whereby two fusion proteins can be produced in said host cell.

45. The host cell of claim 41, further comprising a reporter gene, wherein the expression of the reporter gene is determined by the interaction between the first protein and the second protein.

46. A method for screening for drug candidates capable of modulating the interaction of the proteins of a protein complex, the protein complex selected from the group consisting of the protein complexes of claim 1, said method comprising

(i) combining the proteins of said protein complex in the presence of a drug to form a first complex;

(ii) combining the proteins in the absence of said drug to form a second complex;

(iii) measuring the amount of said first complex and said second complex; and

(iv) comparing the amount of said first complex with the amount of said second complex,

wherein if the amount of said first complex is greater than, or less than the amount of said second complex, then the drug is a drug candidate for modulating the interaction of the proteins of said protein complex.

47. The method of claim 46, wherein said screening is an *in vitro* screening.

48. The method of claim 46, wherein said complex is measured by binding with an antibody specific for said protein complexes.

49. The method of claim 46, wherein if the amount of said first complex is greater than the amount of said second complex, then said drug is a drug candidate for promoting the interaction of said proteins.

50. The method of claim 46, wherein if the amount of said first complex is less than the amount of said second complex, then said drug is a drug candidate for inhibiting the interaction of said proteins.

51. A drug useful for treating a physiological disorder identified by the method of claim 46.

52. The drug of claim 51, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

53. A method of screening for drug candidates useful in treating a physiological disorder which comprises the steps of:

(a) measuring the activity of a protein selected from the goup consisting of a first protein and a second protein in the presence of a drug, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1,

(b) measuring the activity of said protein in the absence of said drug, and

(c) comparing the activity measured in steps (1) and (2),

wherein if there is a difference in activity, then said drug is a drug candidate for treating said physiological disorder.

54. A drug useful for treating a physiological disorder identified by the method of claim 53.

55. The drug of claim 54, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

56. A method for selecting modulators of a protein complex formed between a first protein or a homologue or derivative or fragment thereof and a second protein or a homologue or derivative or fragment thereof, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

providing the protein complex;

contacting said protein complex with a test compound; and

determining the presence or absence of binding of said test compound to said protein complex.

57. A modulator useful for treating a physiological disorder identified by the method of claim 56.

58. The modulator of claim 57, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

59. A method for selecting modulators of an interaction between a first protein and a second protein, said first protein or a homologue or derivative or fragment thereof and said second protein or a homologue or derivative or fragment thereof, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

contacting said first protein with said second protein in the presence of a test compound; and

determining the interaction between said first protein and said second protein.

60. The method of claim 59, wherein at least one of said first and second proteins is a fusion protein having a detectable tag.

61. The method of claim 59, wherein said step of determining the interaction between said first protein and said second protein is conducted in a substantially cell free environment.

62. The method of claim 59, wherein the interaction between said first protein and said second protein is determined in a host cell.

63. The method of claim 62, wherein said host cell is a yeast cell.

64. The method of claim 59, wherein said test compound is provided in a phage display library.

65. The method of claim 59, wherein said test compound is provided in a combinatorial library.

66. A modulator useful for treating a physiological disorder identified by the method of claim 59.

67. The modulator of claim 66, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

68. A method for selecting modulators of a protein complex formed from a first protein or a homologue or derivative or fragment thereof, and a second protein or a homologue or derivative or fragment thereof, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

contacting said protein complex with a test compound; and

determining the interaction between said first protein and said second protein.

69. A modulator useful for treating a physiological disorder identified by the method of claim 68.

70. The modulator of claim 69, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

71. A method for selecting modulators of an interaction between a first polypeptide and a second polypeptide, said first polypeptide being a first protein or a homologue or derivative or fragment thereof and said second polypeptide being a second protein or a homologue or derivative or fragment thereof, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

providing in a host cell a first fusion protein having said first polypeptide, and a second fusion protein having said second polypeptide, wherein a DNA binding domain is fused to one of said first and second polypeptides while a transcription-activating domain is fused to the other of said first and second polypeptides;

providing in said host cell a reporter gene, wherein the transcription of the reporter gene is determined by the interaction between the first polypeptide and the second polypeptide;

allowing said first and second fusion proteins to interact with each other within said host cell in the presence of a test compound; and

determining the presence or absence of expression of said reporter gene.

72. The method of claim 71, wherein said host cell is a yeast cell.

73. A modulator useful for treating a physiological disorder identified by the method of claim 71.

74. The modulator of claim 73, wherein said physiological disorder is selected from the group consisting disorders associated with cholesterol homeostasis and atherogenesis.

75. A method for identifying a compound that binds to a protein in vitro, wherein said protein is selected from the group consisting of the proteins of claim 1, said method comprising:

contacting a test compound with said protein for a time sufficient to form a complex and

detecting for the formation of a complex by detecting said protein or the compound in the complex,

so that if a complex is detected, a compound that binds to protein is identified.

76. A compound useful for treating a physiological disorder identified by the method of claim 75.

77. The compound of claim 76, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

78. A method for selecting modulators of an interaction between a first polypeptide and a second polypeptide, said first polypeptide being a first protein or a homologue or derivative or fragment thereof and said second polypeptide being a second protein or a homologue or derivative or fragment thereof, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

providing atomic coordinates defining a three-dimensional structure of a protein complex formed by said first polypeptide and said second polypeptide; and

designing or selecting compounds capable of modulating the interaction between a first polypeptide and a second polypeptide based on said atomic coordinates.

79. A modulator useful for treating a physiological disorder identified by the method of claim 78.

80. The modulator of claim 79, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

81. A method for providing inhibitors of an interaction between a first polypeptide and a second polypeptide, said first polypeptide being a first protein or a homologue or derivative or fragment thereof and said second polypeptide being a second protein or a homologue or derivative or fragment thereof, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

providing atomic coordinates defining a three-dimensional structure of a protein complex formed by said first polypeptide and said second polypeptide; and

designing or selecting compounds capable of interfering with the interaction between a first polypeptide and a second polypeptide based on said atomic coordinates.

82. An inhibitor useful for treating a physiological disorder identified by the method of claim 81.

83. The inhibitor of claim 82, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

84. A method for selecting modulators of a protein, wherein said protein is selected from the group consisting of the proteins of claim 1, said method comprising:

contacting said protein with a test compound; and

determining binding of said test compound to said protein.

85. The method of claim 84, wherein said test compound is provided in a phage display library.

86. The method of claim 84, wherein said test compound is provided in a combinatorial library.

87. A modulator useful for treating a physiological disorder identified by the method of claim 84.

88. The modulator of claim 87, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

89. A method for modulating, in a cell, a protein complex having a first protein interacting with a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

administering to said cell a compound capable of modulating said protein complex.

90. The method of claim 89, wherein said compound is selected from the group consisting of:

(a) a compound which is capable of interfering with the interaction between said first protein and said second protein,

(b) a compound which is capable of binding at least one of said first protein and said second protein,

(c) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of siad second protein and capable of binding said first protein,

(d) a compound which comprises a peptide capable of binding said first protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of said second protein of the same length,

(e) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of said first protein and capable of binding said second protein,

(f) a compound which comprises a peptide capable of binding said second protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of said first protein of the same length,

(g) a compound which is an antibody immunoreactive with said first protein or said second protein,

(h) a compound which is a nucleic acid encoding an antibody immunoreactive with said first protein or said second protein,

(i) a compound which modulates the expression of said first protein or said second protein,

(j) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding said first protein or complement thereof, and

(k) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding said second protein or complement thereof.

91. A method for modulating, in a cell, a protein complex having a first protein interacting with a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, said method comprising:

administering to said cell a peptide capable of interfering with the interaction between said first protein and said second protein, wherein said peptide is associated with a transporter capable of increasing cellular uptake of said peptide.

92. The method of claim 91, wherein said peptide is covalently linked to said transporter which is selected from the group consisting of penetratins, *l*-$Tat_{49-57}$, *d*-$Tat_{49-57}$, retro-inverso isomers of *l*- or *d*-$Tat_{49-57}$, L-arginine oligomers, D- arginine oligomers, L-lysine oligomers, D-lysine oligomers, L-histine oligomers, D-histine oligomers, L-ornithine oligomers, D-ornithine oligomers, short peptide sequences derived from fibroblast growth factor, Galparan, and HSV-1 structural protein VP22, and peptoid analogs thereof.

93. A method for modulating, in a cell, the interaction of a protein with a ligand, wherein said protein is selected from the group consisting of the first or second proteins of claim 1, said method comprising:

administering to said cell a compound capable of modulating said interaction.

94. The method of claim 93, wherein said protein is one of said first or second proteins and said ligand is the other of said first or second proteins

95. The method of claim 93, wherein said compound is selected from the group consisting of:

(a) a compound which interferes with said interaction,

(b) a compound which binds to said protein or said ligand,

(c) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of said protein and capable of binding said ligand,

(d) a compound which comprises a peptide capable of binding said ligand and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of said protein of the same length,

(e) a compound which is an antibody immunoreactive with said protein or said ligand,

(f) a compound which is a nucleic acid encoding an antibody immunoreactive with said ligand or said protein,

(g) a compound which modulates the expression of said protein or said ligand, and

(h) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding said ligand or said protein or complement thereof.

96. A method for modulating neuronal death in a patient having a physiological disorder comprising:

modulating a protein complex having a first protein interacting with a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1.

97. The method of claim 96, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

98. A method for modulating neuronal death in a patient having physiological disorder comprising:

administering to the patient a compound capable of modulating a protein complex having a first protein interacting with a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1.

99. The method of claim 98, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

100. The method of claim 98, wherein said compound is selected from the group consisting of:

(a) a compound which is capable of interfering with the interaction between said first protein and said second protein,

(b) a compound which is capable of binding at least one of said first protein and said second protein,

(c) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of a second protein and capable of binding a first protein,

(d) a compound which comprises a peptide capable of binding a first protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of a second protein of the same length,

(e) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of first protein and capable of binding a second protein,

(f) a compound which comprises a peptide capable of binding a second protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of a first protein of the same length,

(g) a compound which is an antibody immunoreactive with a first protein or a second protein,

(h) a compound which is a nucleic acid encoding an antibody immunoreactive with a first protein or a second protein,

(i) a compound which modulates the expression of a first protein or a second protein,

(j) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding a first protein or complement thereof, and

(j) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding a second protein or complement thereof

101. A method for modulating neuronal death in a patient having physiological disorder comprising:

administering to said cell a peptide capable of interfering with the interaction between a first protein and a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, wherein said peptide is associated with a transporter capable of increasing cellular uptake of said peptide.

102. The method of claim 101, wherein said peptide is covalently linked to said transporter which is selected from the group consisting of penetratins, *l*-$Tat_{49-57}$, *d*-$Tat_{49-57}$, retro-inverso isomers of *l*- or *d*-$Tat_{49-57}$, L-arginine oligomers, D- arginine oligomers, L-lysine oligomers, D-lysine oligomers, L-histine oligomers, D-histine oligomers, L-ornithine oligomers, D-ornithine oligomers, short peptide sequences derived from fibroblast growth factor, Galparan, and HSV-1 structural protein VP22, and peptoid analogs thereof.

103. A method for treating a physiological disorder comprising:

administering to a patient in need of treatment a compound capable of modulating a protein complex having a first protein interacting with a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1.

104. The method of claim 103, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

105. The method of claim 103, wherein said compound is selected from the group consisting of:

(a) a compound which is capable of interfering with the interaction between said first protein and said second protein,

(b) a compound which is capable of binding at least one of said first protein and said second protein,

(c) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of said second protein and capable of binding said first protein,

(d) a compound which comprises a peptide capable of binding said first protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of said second protein of the same length,

(e) a compound which comprises a peptide having a contiguous span of amino acids of at least 4 amino acids of first protein and capable of binding said second protein,

(f) a compound which comprises a peptide capable of binding said second protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of said first protein of the same length,

(g) a compound which is an antibody immunoreactive with siad first protein or said second protein,

(h) a compound which is a nucleic acid encoding an antibody immunoreactive with said first protein or said second protein,

(i) a compound which modulates the expression of said first protein or said second protein,

(j) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding a first protein or complement thereof,

(k) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding a second protein or complement thereof, and

(l) a compound which is capable of strengthening the interaction between said first protein and said second protein.

106. A method for treating a physiological disorder comprising:

administering to said cell a peptide capable of interfering with the interaction between a first protein and a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1, wherein said peptide is associated with a transporter capable of increasing cellular uptake of said peptide.

107. The method of claim 106, wherein said peptide is covalently linked to said transporter which is selected from the group consisting of penetratins, *l*-$Tat_{49-57}$, *d*-$Tat_{49-57}$, retro-inverso isomers of *l*- or *d*-$Tat_{49-57}$, L-arginine oligomers, D- arginine oligomers, L-lysine oligomers, D-lysine oligomers, L-histine oligomers, D-histine oligomers, L-ornithine oligomers, D-ornithine oligomers, short peptide sequences derived from fibroblast growth factor, Galparan, and HSV-1 structural protein VP22, and peptoid analogs thereof.

108. The method of claim 106, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

109. A method for treating a physiological disorder comprising:

administering to a patient in need of treatment a compound capable of modulating the activity of a first protein or a second protein, wherein said first and second proteins are selected from the group consisting of the proteins of claim 1.

110. The method of claim 109, wherein said physiological disorder is selected from the group consisting of disorders associated with cholesterol homeostasis and atherogenesis.

111. The method of claim 109, wherein the activity is the interaction of said first protein or said second protein with a ligand.

112. The method of claim 111, wherein said ligand is the other of said first or second protein.

113. A method of modulating activity in a cell of a protein, said protein being first protein or a second protein selected from the group consisting of the proteins of claim 1, said method comprising:

administering to said cell a compound capable of modulating said protein.

114. The method of claim 113, wherein said compound is selected from the group consisting of:

(a) a compound which is capable of binding said protein,

(b) a compound which comprises a peptide having a contiguous span of at least 4 amino acids of a first protein and capable of binding a second protein,

(c) a compound which comprises a peptide capable of binding a second protein and having an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of a first protein of the same length,

(d) a compound which is an antibody immunoreactive with said protein,

(e) a compound which is a nucleic acid encoding an antibody immunoreactive with said protein, and

(f) a compound which is an antisense compound or a ribozyme specifically hybridizing to a nucleic acid encoding said protein or complement thereof.

115. A method for modulating activities of a protein in a cell, said protein being a first protein or a second protein selected from the group consisting of the proteins of claim 1, said method comprising:

administering to said cell a peptide having a contiguous span of at least 4 amino acids of one of said first or second proteins and capable of binding the other of said first or second proteins, wherein said peptide is associated with a transporter capable of increasing cellular uptake of said peptide.

116. The method of claim 115, wherein said peptide is covalently linked to said transporter which is selected from the group consisting of penetratins, *l*-$Tat_{49-57}$, *d*-$Tat_{49-57}$, retro-inverso isomers of *l*- or *d*-$Tat_{49-57}$, L-arginine oligomers, D- arginine oligomers, L-lysine oligomers, D-lysine oligomers, L-histine oligomers, D-histine oligomers, L-ornithine oligomers, D-ornithine oligomers, short peptide sequences derived from fibroblast growth factor, Galparan, and HSV-1 structural protein VP22, and peptoid analogs thereof.

117. An isolated nucleic acid encoding a protein comprising an amino acid sequence set forth in SEQ ID NO:4.

118. The isolated nucleic acid sequence of claim 117 which comprises nucleotides 544-6960 of SEQ ID NO:3 or complement thereof.

119. An isolated nucleic acid encoding a protein comprising an amino acid sequence which is at least 70% identical to the amino acid sequence set forth in SEQ ID NO:4 and which is capable of interacting with LXR-alpha.

120. An isolated nucleic acid comprising a nucleotide sequence which is at least 60% identical to nucleotides 544-6960 of SEQ ID NO:3 or complement thereof.

121. An isolated nucleic acid sequence comprising a nucleotide sequence set forth in SEQ ID NO:3 or complement thereof.

122. An isolated nucleic acid comprising a contiguous span of at least 17 nucleotides of the nucleotide sequence set forth in SEQ ID NO:3 or complement thereof.

123. The isolated nucleic acid of claim 122 comprising at least 21 nucleotides.

124. The isolated nucleic acid of claim 122 comprising at least 25 nucleotides.

125. The isolated nucleic acid of claim 122 comprising at least 30 nucleotides.

126. The isolated nucleic acid of claim 122 comprising at least 50 nucleotides.

127. An isolated nucleic acid comprising at least 21 nucleotides that encodes a contiguous span of at least 7 amino acids of the amino acid sequence set forth in SEQ ID NO:4.

128. The isolated nucleic acid of claim 127 encoding at least 8 contiguous amino acids.

129. The isolated nucleic acid of claim 127 encoding at least 9 contiguous amino acids.

130. The isolated nucleic acid of claim 127 encoding at least 10 contiguous amino acids.

131. The isolated nucleic acid of claim 127 encoding at least 15 contiguous amino acids.

132. The isolated nucleic acid of claim 127 encoding at least 20 contiguous amino acids.

133. The isolated nucleic acid of claim 127 encoding at least 25 contiguous amino acids.

134. A nucleic acid vector comprising the isolated nucleic acid of claim 117.

135. A nucleic acid vector comprising the isolated nucleic acid of claim 118.

136. A nucleic acid vector comprising the isolated nucleic acid of claim 119.

137. A nucleic acid vector comprising the isolated nucleic acid of claim 124.

138. A nucleic acid vector comprising the isolated nucleic acid of claim 130.

139. A host cell comprising the isolated nucleic acid of claim 117.

140. A host cell comprising the isolated nucleic acid of claim 118.

141. A host cell comprising the isolated nucleic acid of claim 119.

142. A host cell comprising the isolated nucleic acid of claim 116.

143. A host cell comprising the isolated nucleic acid of claim 130.

144. A microarray comprising the isolated nucleic acid of claim 130.

145. An isolated polypeptide comprising an amino acid sequence set forth in SEQ ID NO:4.

146. An isolated polypeptide comprising an amino acid sequence that is at least 70% identical to the amino acid sequence set forth in SEQ ID NO:4 and capable of interacting with LXR-alpha.

147. An isolated polypeptide comprising a contiguous span of at least 8 amino acids of the amino acid sequence set forth in SEQ ID NO:4.

148. The isolated polypeptide of claim 147 comprising a contiguous span of at least 10 amino acids.

149. The isolated polypeptide of claim 147 comprising a contiguous span of at least 12 amino acids.

150. The isolated polypeptide of claim 147 comprising a contiguous span of at least 15 amino acids.

151. The isolated polypeptide of claim 147 comprising a contiguous span of at least 17 amino acids.

152. The isolated polypeptide of claim 147 comprising a contiguous span of at least 20 amino acids.

153. An isolated polypeptide comprising an amino acid sequence of from 4 to 30 amino acids that is at least 75% identical to a contiguous span of amino acids of the amino acid sequence set forth in SEQ ID NO:4 of the same length, wherein said isolated polypeptide is capable of interacting with LXR-alpha.

154. The isolated polypeptide of claim 153, wherein said amino acid sequence comprises from 8 to 20 amino acids.

155. An antibody which is specifically immunoreactive with the isolated polypeptide of claim 145.

156. An antibody which is specifically immunoreactive with the isolated polypeptide of claim 147.

157. A protein microarray comprising the isolated polypeptide of claim 145.

158. A protein microarray comprising the isolated polypeptide of claim 147.

159. A protein microarray comprising the isolated polypeptide of claim 154.

160. A method for making an isolated polypeptide comprising an amino acid sequence set forth in SEQ ID NO:4, comprising:

providing an expression vector comprising a nucleic acid encoding said amino acid sequence; and

introducing said expression vector into a host cell such that said host cell producing the isolated polypeptide.

5

SEQUENCE LISTING

<110> Myriad Genetics, Inc.
Cimbora, Daniel M.
Heichman, Karen
Bartel, Paul L.

<120> Protein-Protein Interactions

<130> 2318-282-II

<150> US 60/256,983
<151> 2000-12-21

<160> 4

<170> PatentIn version 3.0

<210> 1
<211> 40
<212> DNA
<213> Artificial

<220>
<223> oligonucleotide primer

<400> 1
gcaggaaaca gctatgacca tacagtcagc ggccgccacc 40

<210> 2
<211> 39
<212> DNA
<213> Artificial

<220>
<223> oligonucleotide primer

<400> 2
acggccagtc gcgtggagtg ttatgtcatg cggccgcta 39

<210> 3
<211> 10625
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (544)..(6960)

<400> 3
cttattttga aaacatttac atagtgatta gttaacccaa cagaccaatc ctgggaagac 60

agccagagcc tgcagcacct tagtaacaga aaaactgata attaggagaa gagacctgtc 120

caagaccagg aacctggacc aaaattgtgc catgttgctt tactttaatg agtggcccca 180

gtaaaaactg agctgtatgg cagagctgtt cacatttatc ttctgtgtcc acccagttct 240

gctgaaaccc ctggcaagat cgtggccctg ttgtagcttg tcatgttttg aacagctgtc 300

tatggaaaga aagcaaacac aacctagagc aacattgatt tgttttagaa agctctttta 360
ttttcagttc tggctgtgtt caacatctta gcttacgttt ttcatgttgt aatgatctgc 420

cgtatggacg atcacctcta agttagagag ttctgtaatt tggcttggat taaagatgct 480

```
tggttagtga aagctgctgc ttttttttata gtcaaaggac tggttctgag agccttgttg    540

cag atg gct gag gtc acc gtc cca agg gtg tat gtc gtg ttt ggc atc      588
    Met Ala Glu Val Thr Val Pro Arg Val Tyr Val Val Phe Gly Ile
    1               5                   10                  15

cat tgc atc atg gcg aag gca tct tca gat gtg cag gtt tca ggc ttt      636
His Cys Ile Met Ala Lys Ala Ser Ser Asp Val Gln Val Ser Gly Phe
                20                  25                  30

cat cgg aaa atc cag cac gtt aaa aat gaa ctt tgc cac atg ttg agc      684
His Arg Lys Ile Gln His Val Lys Asn Glu Leu Cys His Met Leu Ser
            35                  40                  45

ttg gag gag gtg gcc cca gtg ctg cag cag aca tta ctt cag gac aac      732
Leu Glu Glu Val Ala Pro Val Leu Gln Gln Thr Leu Leu Gln Asp Asn
        50                  55                  60

ctc ttg ggc agg gta cat ttt gac caa ttt aaa gaa gca tta ata ctc      780
Leu Leu Gly Arg Val His Phe Asp Gln Phe Lys Glu Ala Leu Ile Leu
    65                  70                  75

atc ttg tcc aga act ctg tca aat gaa gaa cac ttt caa gaa cca gac      828
Ile Leu Ser Arg Thr Leu Ser Asn Glu Glu His Phe Gln Glu Pro Asp
80                  85                  90                  95

tgc tca cta gaa gct cag ccc aaa tat gtt aga ggt ggg aag cgt tac      876
Cys Ser Leu Glu Ala Gln Pro Lys Tyr Val Arg Gly Gly Lys Arg Tyr
                100                 105                 110

gga cga agg tcc ttg ccc gag ttc caa gag tcc gtg gag gag ttt cct      924
Gly Arg Arg Ser Leu Pro Glu Phe Gln Glu Ser Val Glu Glu Phe Pro
            115                 120                 125

gaa gtg acg gtg att gag cca ctg gat gaa gaa gcg cgg cct tca cac      972
Glu Val Thr Val Ile Glu Pro Leu Asp Glu Glu Ala Arg Pro Ser His
        130                 135                 140

atc cca gcc ggt gac tgc agt gag cac tgg aag acg caa cgc agt gag     1020
Ile Pro Ala Gly Asp Cys Ser Glu His Trp Lys Thr Gln Arg Ser Glu
    145                 150                 155

gag tat gaa gcg gaa ggc cag tta agg ttt tgg aac cca gat gac ttg     1068
Glu Tyr Glu Ala Glu Gly Gln Leu Arg Phe Trp Asn Pro Asp Asp Leu
160                 165                 170                 175

aat gct tca cag agt gga tct tcc cct ccc caa gac tgg ata gaa gag     1116
Asn Ala Ser Gln Ser Gly Ser Ser Pro Pro Gln Asp Trp Ile Glu Glu
                180                 185                 190

aaa ctg caa gaa gtt tgt gaa gat ttg ggg atc acc cgt gat ggt cac     1164
Lys Leu Gln Glu Val Cys Glu Asp Leu Gly Ile Thr Arg Asp Gly His
            195                 200                 205

ctg aac cgg aag aag ctg gtc tcc atc tgt gag cag tat ggt tta cag     1212
Leu Asn Arg Lys Lys Leu Val Ser Ile Cys Glu Gln Tyr Gly Leu Gln
        210                 215                 220
aat gtg gat gga gag atg ctc gag gaa gta ttc cat aat ctt gat cct     1260
Asn Val Asp Gly Glu Met Leu Glu Glu Val Phe His Asn Leu Asp Pro
    225                 230                 235

gac ggt aca atg agt gta gaa gat ttt ttc tat ggt ttg ttt aaa aat     1308
Asp Gly Thr Met Ser Val Glu Asp Phe Phe Tyr Gly Leu Phe Lys Asn
```

```
240                 245                 250                 255

gga aaa tct ctt aca cca tca gca tct act cca tat aga caa cta aaa     1356
Gly Lys Ser Leu Thr Pro Ser Ala Ser Thr Pro Tyr Arg Gln Leu Lys
                260                 265                 270

agg cac ctt tcc atg cag tct ttc gat gag agt gga cga cgt acc aca     1404
Arg His Leu Ser Met Gln Ser Phe Asp Glu Ser Gly Arg Arg Thr Thr
            275                 280                 285

acc tca tca gca atg aca agt acc att ggc ttt cgg gtc ttc tcc tgc     1452
Thr Ser Ser Ala Met Thr Ser Thr Ile Gly Phe Arg Val Phe Ser Cys
        290                 295                 300

ctg gat gat ggg atg ggc cat gca tct gtg gag aga ata ctg gac acc     1500
Leu Asp Asp Gly Met Gly His Ala Ser Val Glu Arg Ile Leu Asp Thr
    305                 310                 315

tgg cag gaa gag ggc att gag aac agc cag gag atc ctg aag gcc ttg     1548
Trp Gln Glu Glu Gly Ile Glu Asn Ser Gln Glu Ile Leu Lys Ala Leu
320                 325                 330                 335

gat ttc agc ctc gat gga aac atc aat ttg aca gaa tta aca ctg gcc     1596
Asp Phe Ser Leu Asp Gly Asn Ile Asn Leu Thr Glu Leu Thr Leu Ala
                340                 345                 350

ctt gaa aat gaa ctt ttg gtt acc aag aac agc att cac cag gcg gct     1644
Leu Glu Asn Glu Leu Leu Val Thr Lys Asn Ser Ile His Gln Ala Ala
            355                 360                 365

ctg gcc agc ttt aag gct gaa atc cgg cat ttg ttg gaa cga gtt gat     1692
Leu Ala Ser Phe Lys Ala Glu Ile Arg His Leu Leu Glu Arg Val Asp
        370                 375                 380

cag gtg gtc aga gaa aaa gag aag cta cgg tca gat ctg gac aag gcc     1740
Gln Val Val Arg Glu Lys Glu Lys Leu Arg Ser Asp Leu Asp Lys Ala
    385                 390                 395

gag aag ctc aag tct tta atg gcc tcg gag gtg gat gat cac cat gcg     1788
Glu Lys Leu Lys Ser Leu Met Ala Ser Glu Val Asp Asp His His Ala
400                 405                 410                 415

gcc ata gag cgg cgg aat gag tac aac ctc agg aaa ctg gat gga gag     1836
Ala Ile Glu Arg Arg Asn Glu Tyr Asn Leu Arg Lys Leu Asp Gly Glu
                420                 425                 430

tac aag gag cga ata gca gcc tta aaa aat gaa ctc cga aaa gag aga     1884
Tyr Lys Glu Arg Ile Ala Ala Leu Lys Asn Glu Leu Arg Lys Glu Arg
            435                 440                 445

gag cag atc ctg cag cag gca ggc aag cag cgt tta gaa ctt gaa cag     1932
Glu Gln Ile Leu Gln Gln Ala Gly Lys Gln Arg Leu Glu Leu Glu Gln
        450                 455                 460

gaa att gaa aag gca aaa aca gaa gag aac tat atc cgg gac cgc ctt     1980
Glu Ile Glu Lys Ala Lys Thr Glu Glu Asn Tyr Ile Arg Asp Arg Leu
    465                 470                 475

gcc ctc tct tta aag gaa aac agt cgt ctg gaa aat gag ctt cta gaa     2028
Ala Leu Ser Leu Lys Glu Asn Ser Arg Leu Glu Asn Glu Leu Leu Glu
480                 485                 490                 495

aat gca gag aag ttg gca gaa tat gag aat ctg aca aac aaa ctt cag     2076
Asn Ala Glu Lys Leu Ala Glu Tyr Glu Asn Leu Thr Asn Lys Leu Gln
```

```
                500                 505                 510

aga aat ttg gaa aat gtg tta gca gaa aag ttt ggt gac ctc gat cct     2124
Arg Asn Leu Glu Asn Val Leu Ala Glu Lys Phe Gly Asp Leu Asp Pro
            515                 520                 525

agc agt gct gag ttc ttc ctg caa gaa gag aga ctg aca cag atg aga     2172
Ser Ser Ala Glu Phe Phe Leu Gln Glu Glu Arg Leu Thr Gln Met Arg
        530                 535                 540

aat gaa tat gag cgg cag tgc agg gta cta caa gac caa gta gat gaa     2220
Asn Glu Tyr Glu Arg Gln Cys Arg Val Leu Gln Asp Gln Val Asp Glu
    545                 550                 555

ctc cag tct gag ctg gaa gaa tat cgt gca caa ggc aga gtg ctc agg     2268
Leu Gln Ser Glu Leu Glu Glu Tyr Arg Ala Gln Gly Arg Val Leu Arg
560                 565                 570                 575

ctt ccg ttg aag aac tca ccg tca gaa gaa gtt gag gct aac agc ggt     2316
Leu Pro Leu Lys Asn Ser Pro Ser Glu Glu Val Glu Ala Asn Ser Gly
                580                 585                 590

ggc att gag ccc gaa cac ggg ctc ggt tct gaa gaa tgc aat cca ttg     2364
Gly Ile Glu Pro Glu His Gly Leu Gly Ser Glu Glu Cys Asn Pro Leu
            595                 600                 605

aat atg agc att gag gca gag ctg gtc att gaa cag atg aaa gaa caa     2412
Asn Met Ser Ile Glu Ala Glu Leu Val Ile Glu Gln Met Lys Glu Gln
        610                 615                 620

cat cac agg gac ata tgt tgc ctc aga ctg gag ctc gaa gat aaa gtg     2460
His His Arg Asp Ile Cys Cys Leu Arg Leu Glu Leu Glu Asp Lys Val
    625                 630                 635
__s20
cgc cat tat gaa aag cag ctg gac gaa acc gtg gtc agc tgc aag aag     2508
Arg His Tyr Glu Lys Gln Leu Asp Glu Thr Val Val Ser Cys Lys Lys
640                 645                 650                 655

gca cag gag aac atg aag caa agg cat gag aac gaa acg cgc acc tta     2556
Ala Gln Glu Asn Met Lys Gln Arg His Glu Asn Glu Thr Arg Thr Leu
                660                 665                 670

gaa aaa caa ata agt gac ctt aaa aat gaa att gct gaa ctt cag ggg     2604
Glu Lys Gln Ile Ser Asp Leu Lys Asn Glu Ile Ala Glu Leu Gln Gly
            675                 680                 685

caa gca gca gtg ctc aag gag gca cat cat gag gcc act tgc agg cat     2652
Gln Ala Ala Val Leu Lys Glu Ala His His Glu Ala Thr Cys Arg His
        690                 695                 700

gag gag gag aaa aaa caa ctg caa gtg aag ctt gag gag gaa aag act     2700
Glu Glu Glu Lys Lys Gln Leu Gln Val Lys Leu Glu Glu Glu Lys Thr
    705                 710                 715

cac ctg cag gag aag ctg agg ctg caa cat gag atg gag ctc aag gct     2748
His Leu Gln Glu Lys Leu Arg Leu Gln His Glu Met Glu Leu Lys Ala
720                 725                 730                 735

aga ctg aca cag gct caa gca agc ttt gag cgg gag agg gaa ggc ctt     2796
Arg Leu Thr Gln Ala Gln Ala Ser Phe Glu Arg Glu Arg Glu Gly Leu
                740                 745                 750

cag agt agc gcc tgg aca gaa gag aag gtg aga ggc ttg act cag gaa     2844
Gln Ser Ser Ala Trp Thr Glu Glu Lys Val Arg Gly Leu Thr Gln Glu
```

```
            755                 760                 765

cta gag cag ttt cac cag gag cag ctg aca agc ctg gtg gag aaa cac     2892
Leu Glu Gln Phe His Gln Glu Gln Leu Thr Ser Leu Val Glu Lys His
        770                 775                 780

act ctt gag aaa gag gag tta aga aaa gag ctc ttg gaa aag cac caa     2940
Thr Leu Glu Lys Glu Glu Leu Arg Lys Glu Leu Leu Glu Lys His Gln
    785                 790                 795

agg gag ctt cag gag gga agg gaa aaa atg gaa aca gag tgt aat aga     2988
Arg Glu Leu Gln Glu Gly Arg Glu Lys Met Glu Thr Glu Cys Asn Arg
800                 805                 810                 815

aga acc tct caa ata gaa gcc cag ttt cag tct gat tgt cag aaa gtc     3036
Arg Thr Ser Gln Ile Glu Ala Gln Phe Gln Ser Asp Cys Gln Lys Val
                820                 825                 830

act gag agg tgt gaa agc gct ctg caa agc ctg gag ggg cgc tac cgc     3084
Thr Glu Arg Cys Glu Ser Ala Leu Gln Ser Leu Glu Gly Arg Tyr Arg
            835                 840                 845

caa gag ctg aag gac ctc cag gaa cag cag cgt gag gag aaa tcc cag     3132
Gln Glu Leu Lys Asp Leu Gln Glu Gln Gln Arg Glu Glu Lys Ser Gln
        850                 855                 860

tgg gaa ttt gag aag gac gag ctc acc cag gag tgt gcg gaa gcc cag     3180
Trp Glu Phe Glu Lys Asp Glu Leu Thr Gln Glu Cys Ala Glu Ala Gln
    865                 870                 875

gag ctg ctg aaa gag act ctt aag aga gag aaa aca act tct ctg gtc     3228
Glu Leu Leu Lys Glu Thr Leu Lys Arg Glu Lys Thr Thr Ser Leu Val
880                 885                 890                 895

ctg acc cag gag aga gag atg ctg gag aaa aca tac aaa gaa cat ttg     3276
Leu Thr Gln Glu Arg Glu Met Leu Glu Lys Thr Tyr Lys Glu His Leu
                900                 905                 910

aac agc atg gtc gtc gag aga cag cag cta ctc caa gac ctg gaa gac     3324
Asn Ser Met Val Val Glu Arg Gln Gln Leu Leu Gln Asp Leu Glu Asp
            915                 920                 925

cta aga aat gta tct gaa acc cag caa agc ctg ctg tct gac cag ata     3372
Leu Arg Asn Val Ser Glu Thr Gln Gln Ser Leu Leu Ser Asp Gln Ile
        930                 935                 940

ctt gag ctg aag agc agt cac aaa agg gaa ctg agg gag cgt gag gag     3420
Leu Glu Leu Lys Ser Ser His Lys Arg Glu Leu Arg Glu Arg Glu Glu
    945                 950                 955
gtc ctg tgc cag gca ggg gct tcg gag cag ctg gcc agc cag cgg ctg     3468
Val Leu Cys Gln Ala Gly Ala Ser Glu Gln Leu Ala Ser Gln Arg Leu
960                 965                 970                 975

gaa aga cta gaa atg gaa cat gac cag gaa agg cag gaa atg atg tcc     3516
Glu Arg Leu Glu Met Glu His Asp Gln Glu Arg Gln Glu Met Met Ser
                980                 985                 990

aag ctt cta gcc atg gag aac att cac  aaa gcg acc tgt gag  aca gca   3564
Lys Leu Leu Ala Met Glu Asn Ile His  Lys Ala Thr Cys Glu  Thr Ala
            995                 1000                1005

gat cga gaa  aga gcc gag atg agc  aca gaa atc tcc aga  ctt cag      3609
Asp Arg Glu  Arg Ala Glu Met Ser  Thr Glu Ile Ser Arg  Leu Gln
        1010                1015                1020
```

```
agt aaa ata  aag gaa atg cag cag  gca aca tct cct ctc  tca atg      3654
Ser Lys Ile  Lys Glu Met Gln Gln  Ala Thr Ser Pro Leu  Ser Met
        1025                 1030                 1035

ctt cag agt  ggt tgc cag gtg ata  gga gag gag gag gtg  gaa gga      3699
Leu Gln Ser  Gly Cys Gln Val Ile  Gly Glu Glu Glu Val  Glu Gly
        1040                 1045                 1050

gat gga gcc  ctg tcc ctg ctt cag  caa ggg gag cag ctg  ttg gaa      3744
Asp Gly Ala  Leu Ser Leu Leu Gln  Gln Gly Glu Gln Leu  Leu Glu
        1055                 1060                 1065

gaa aat ggg  gac gtc ctc tta agc  ctg cag aga gct cat  gaa cag      3789
Glu Asn Gly  Asp Val Leu Leu Ser  Leu Gln Arg Ala His  Glu Gln
        1070                 1075                 1080

gca gtg aag  gaa aat gtg aaa atg  gct act gaa att tct  aga ttg      3834
Ala Val Lys  Glu Asn Val Lys Met  Ala Thr Glu Ile Ser  Arg Leu
        1085                 1090                 1095

caa cag agg  cta caa aag tta gag  cca ggg tta gta atg  tct tct      3879
Gln Gln Arg  Leu Gln Lys Leu Glu  Pro Gly Leu Val Met  Ser Ser
        1100                 1105                 1110

tgt ttg gat  gag cca gct act gag  ttt ttt gga aat act  gcg gaa      3924
Cys Leu Asp  Glu Pro Ala Thr Glu  Phe Phe Gly Asn Thr  Ala Glu
        1115                 1120                 1125

caa aca gag  cag ttt tta cag caa  aac cga acg aag caa  gta gaa      3969
Gln Thr Glu  Gln Phe Leu Gln Gln  Asn Arg Thr Lys Gln  Val Glu
        1130                 1135                 1140

ggt gtg acc  agg cgg cat gtc cta  agt gac ctg gaa gat  gat gag      4014
Gly Val Thr  Arg Arg His Val Leu  Ser Asp Leu Glu Asp  Asp Glu
        1145                 1150                 1155

gtc cgg gac  ctg gga agt aca ggg  acg agc tct gtt cag  aga cag      4059
Val Arg Asp  Leu Gly Ser Thr Gly  Thr Ser Ser Val Gln  Arg Gln
        1160                 1165                 1170

gaa gtc aaa  ata gag gag tct gaa  gct tca gta gag ggt  ttt tct      4104
Glu Val Lys  Ile Glu Glu Ser Glu  Ala Ser Val Glu Gly  Phe Ser
        1175                 1180                 1185

gag ctt gaa  aac agt gaa gag acc  agg act gaa tcc tgg  gag ctg      4149
Glu Leu Glu  Asn Ser Glu Glu Thr  Arg Thr Glu Ser Trp  Glu Leu
        1190                 1195                 1200

aag aat cag  att agt cag ctt cag  gaa cag cta atg atg  tta tgt      4194
Lys Asn Gln  Ile Ser Gln Leu Gln  Glu Gln Leu Met Met  Leu Cys
        1205                 1210                 1215

gcg gac tgt  gat cga gct tct gaa  aag aaa cag gac cta  ctt ttt      4239
Ala Asp Cys  Asp Arg Ala Ser Glu  Lys Lys Gln Asp Leu  Leu Phe
        1220                 1225                 1230

gat gtt tct  gtg cta aaa aag aaa  ctg aag atg ctt gag  aga atc      4284
Asp Val Ser  Val Leu Lys Lys Lys  Leu Lys Met Leu Glu  Arg Ile
        1235                 1240                 1245

cct gag gct  tct ccc aaa tat aag  ctg ttg tat gaa gat  gtg agc      4329
Pro Glu Ala  Ser Pro Lys Tyr Lys  Leu Leu Tyr Glu Asp  Val Ser
        1250                 1255                 1260
```

```
cga gaa aat  gac tgc ctt cag gaa  gag ctg aga atg atg  gag aca      4374
Arg Glu Asn  Asp Cys Leu Gln Glu  Glu Leu Arg Met Met  Glu Thr
        1265                 1270                 1275

cgc tac gat  gag gca cta gaa aat  aac aaa gaa ctc act  gca gag      4419
Arg Tyr Asp  Glu Ala Leu Glu Asn  Asn Lys Glu Leu Thr  Ala Glu
        1280                 1285                 1290

gtt ttc agg  ttg cag gat gag ctg  aag aaa atg gag gaa  gtc act      4464
Val Phe Arg  Leu Gln Asp Glu Leu  Lys Lys Met Glu Glu  Val Thr
        1295                 1300                 1305

gaa aca ttc  ctc agc ctg gaa aag  agt tac gat gag gtc  aaa ata      4509
Glu Thr Phe  Leu Ser Leu Glu Lys  Ser Tyr Asp Glu Val  Lys Ile
        1310                 1315                 1320

gaa aat gag  ggg ctg aat gtt ctg  gtt ttg aga ctt caa  ggc aag      4554
Glu Asn Glu  Gly Leu Asn Val Leu  Val Leu Arg Leu Gln  Gly Lys
        1325                 1330                 1335

att gag aag  ctt cag gaa agc gtg  gtc cag cgg tgt gac  tgc tgc      4599
Ile Glu Lys  Leu Gln Glu Ser Val  Val Gln Arg Cys Asp  Cys Cys
        1340                 1345                 1350

tta tgg gaa  gcc agt tta gag aac  ctg gaa atc gaa cct  gat gga      4644
Leu Trp Glu  Ala Ser Leu Glu Asn  Leu Glu Ile Glu Pro  Asp Gly
        1355                 1360                 1365

aat ata ctc  cag ctc aat cag aca  ctg gaa gag tgt gtg  ccc agg      4689
Asn Ile Leu  Gln Leu Asn Gln Thr  Leu Glu Glu Cys Val  Pro Arg
        1370                 1375                 1380

gtt agg agt  gta cat cat gtc ata  gag gaa tgt aag caa  gaa aac      4734
Val Arg Ser  Val His His Val Ile  Glu Glu Cys Lys Gln  Glu Asn
        1385                 1390                 1395

cag tac ctt  gag ggg aac aca cag  ctc ttg gaa aaa gta  aaa gca      4779
Gln Tyr Leu  Glu Gly Asn Thr Gln  Leu Leu Glu Lys Val  Lys Ala
        1400                 1405                 1410

cat gaa att  gcc tgg tta cat gga  aca att cag aca cat  caa gaa      4824
His Glu Ile  Ala Trp Leu His Gly  Thr Ile Gln Thr His  Gln Glu
        1415                 1420                 1425

agg cca aga  gta cag aat caa gtt  ata ctg gag gaa aac  act act      4869
Arg Pro Arg  Val Gln Asn Gln Val  Ile Leu Glu Glu Asn  Thr Thr
        1430                 1435                 1440

ctc cta ggc  ttt caa gac aaa cat  ttt cag cat cag gcc  acc ata      4914
Leu Leu Gly  Phe Gln Asp Lys His  Phe Gln His Gln Ala  Thr Ile
        1445                 1450                 1455

gca gag tta  gaa ctg gag aaa aca  aag tta cag gag ctg  act agg      4959
Ala Glu Leu  Glu Leu Glu Lys Thr  Lys Leu Gln Glu Leu  Thr Arg
        1460                 1465                 1470

aag ttg aag  gag aga gtc act att  tta gtt aag caa aaa  gat gta      5004
Lys Leu Lys  Glu Arg Val Thr Ile  Leu Val Lys Gln Lys  Asp Val
        1475                 1480                 1485

ctt tct cac  gga gaa aag gag gaa  gag ctg aag gca atg  atg cat      5049
Leu Ser His  Gly Glu Lys Glu Glu  Glu Leu Lys Ala Met  Met His
        1490                 1495                 1500
```

```
gac ttg cag  atc acg tgc agt gag  atg cag caa aaa gtt  gaa ctt      5094
Asp Leu Gln  Ile Thr Cys Ser Glu  Met Gln Gln Lys Val  Glu Leu
        1505                 1510                 1515

ctg aga tat  gaa tct gaa aag ctt  caa cag gaa aat tct  att ttg      5139
Leu Arg Tyr  Glu Ser Glu Lys Leu  Gln Gln Glu Asn Ser  Ile Leu
        1520                 1525                 1530

aga aat gaa  att act act tta aat  gaa gaa gat agc att  tct aac      5184
Arg Asn Glu  Ile Thr Thr Leu Asn  Glu Glu Asp Ser Ile  Ser Asn
        1535                 1540                 1545

ctg aaa tta  ggg aca tta aat gga  tct cag gaa gaa atg  tgg caa      5229
Leu Lys Leu  Gly Thr Leu Asn Gly  Ser Gln Glu Glu Met  Trp Gln
        1550                 1555                 1560

aaa acg gaa  act gta aaa caa gaa  aat gct gca gtt cag  aag atg      5274
Lys Thr Glu  Thr Val Lys Gln Glu  Asn Ala Ala Val Gln  Lys Met
        1565                 1570                 1575

gtt gaa aat  tta aag aaa cag att  tca gaa tta aaa atc  aaa aac      5319
Val Glu Asn  Leu Lys Lys Gln Ile  Ser Glu Leu Lys Ile  Lys Asn
        1580                 1585                 1590

caa caa ttg  gat ttg gaa aat aca  gaa ctt agc caa aag  aac tct      5364
Gln Gln Leu  Asp Leu Glu Asn Thr  Glu Leu Ser Gln Lys  Asn Ser
        1595                 1600                 1605

caa aac cag  gaa aaa ctg caa gaa  ctt aat caa cgt cta  aca gaa      5409
Gln Asn Gln  Glu Lys Leu Gln Glu  Leu Asn Gln Arg Leu  Thr Glu
        1610                 1615                 1620

atg cta tgc  cag aag gaa aaa gag  cca gga aac agt gca  ttg gag      5454
Met Leu Cys  Gln Lys Glu Lys Glu  Pro Gly Asn Ser Ala  Leu Glu
        1625                 1630                 1635

gaa cgg gaa  caa gag aag ttt aat  ctg aaa gaa gaa ctg  gaa cgt      5499
Glu Arg Glu  Gln Glu Lys Phe Asn  Leu Lys Glu Glu Leu  Glu Arg
        1640                 1645                 1650
tgt aaa gtg  cag tcc tcc act tta  gtg tct tct ctg gag  gcg gag      5544
Cys Lys Val  Gln Ser Ser Thr Leu  Val Ser Ser Leu Glu  Ala Glu
        1655                 1660                 1665

ctc tct gaa  gtt aaa ata cag acc  cat att gtg caa cag  gaa aac      5589
Leu Ser Glu  Val Lys Ile Gln Thr  His Ile Val Gln Gln  Glu Asn
        1670                 1675                 1680

cac ctt ctc  aaa gat gaa ctg gag  aaa atg aaa cag ctg  cac aga      5634
His Leu Leu  Lys Asp Glu Leu Glu  Lys Met Lys Gln Leu  His Arg
        1685                 1690                 1695

tgt ccc gat  ctc tct gac ttc cag  caa aaa atc tct agt  gtt cta      5679
Cys Pro Asp  Leu Ser Asp Phe Gln  Gln Lys Ile Ser Ser  Val Leu
        1700                 1705                 1710

agc tac aac  gaa aaa ctg ctg aaa  gaa aag gaa gct ctg  agt gag      5724
Ser Tyr Asn  Glu Lys Leu Leu Lys  Glu Lys Glu Ala Leu  Ser Glu
        1715                 1720                 1725

gaa tta aat  agc tgt gtc gat aag  ttg gca aaa tca agt  ctt tta      5769
Glu Leu Asn  Ser Cys Val Asp Lys  Leu Ala Lys Ser Ser  Leu Leu
        1730                 1735                 1740
```

```
gag cat aga  att gcg acg atg aag  cag gaa cag aaa tcc  tgg gaa      5814
Glu His Arg  Ile Ala Thr Met Lys  Gln Glu Gln Lys Ser  Trp Glu
       1745                 1750                 1755

cat cag agt  gcg agc tta aag tca  cag ctg gtg gct tct  cag gaa      5859
His Gln Ser  Ala Ser Leu Lys Ser  Gln Leu Val Ala Ser  Gln Glu
       1760                 1765                 1770

aag gtt cag  aat tta gaa gac acc  gtg cag aat gta aac  ctg caa      5904
Lys Val Gln  Asn Leu Glu Asp Thr  Val Gln Asn Val Asn  Leu Gln
       1775                 1780                 1785

atg tcc cgg  atg aaa tct gac cta  cga gtg act cag cag  gaa aag      5949
Met Ser Arg  Met Lys Ser Asp Leu  Arg Val Thr Gln Gln  Glu Lys
       1790                 1795                 1800

gag gct tta  aaa caa gaa gtg atg  tct tta cat aag caa  ctt cag      5994
Glu Ala Leu  Lys Gln Glu Val Met  Ser Leu His Lys Gln  Leu Gln
       1805                 1810                 1815

aat gct ggt  ggc aag agc tgg gcc  cca gag ata gct act  cat cca      6039
Asn Ala Gly  Gly Lys Ser Trp Ala  Pro Glu Ile Ala Thr  His Pro
       1820                 1825                 1830

tca ggg ctc  cat aac cag cag aaa  agg ctg tcc tgg gac  aag ttg      6084
Ser Gly Leu  His Asn Gln Gln Lys  Arg Leu Ser Trp Asp  Lys Leu
       1835                 1840                 1845

gat cat ctg  atg aat gag gaa cag  cag ctg ctt tgg caa  gag aat      6129
Asp His Leu  Met Asn Glu Glu Gln  Gln Leu Leu Trp Gln  Glu Asn
       1850                 1855                 1860

gag agg ctc  cag acc atg gta cag  aac acc aaa gcc gaa  ctc acg      6174
Glu Arg Leu  Gln Thr Met Val Gln  Asn Thr Lys Ala Glu  Leu Thr
       1865                 1870                 1875

cac tcc cgg  gag aag gtc cgt caa  ttg gaa tcc aat ctt  ctt ccc      6219
His Ser Arg  Glu Lys Val Arg Gln  Leu Glu Ser Asn Leu  Leu Pro
       1880                 1885                 1890

aag cac caa  aaa cat cta aac cca  tca ggt acc atg aat  ccc aca      6264
Lys His Gln  Lys His Leu Asn Pro  Ser Gly Thr Met Asn  Pro Thr
       1895                 1900                 1905

gag caa gaa  aaa ttg agc tta aag  aga gag tgt gat cag  ttt cag      6309
Glu Gln Glu  Lys Leu Ser Leu Lys  Arg Glu Cys Asp Gln  Phe Gln
       1910                 1915                 1920

aaa gaa caa  tct cct gct aac agg  aag gtc agt cag atg  aat tcc      6354
Lys Glu Gln  Ser Pro Ala Asn Arg  Lys Val Ser Gln Met  Asn Ser
       1925                 1930                 1935

ctt gaa caa  gaa tta gaa aca att  cat ttg gaa aat gaa  ggc ctg      6399
Leu Glu Gln  Glu Leu Glu Thr Ile  His Leu Glu Asn Glu  Gly Leu
       1940                 1945                 1950

aaa aag aaa  caa gta aaa ctg gat  gag cag ctc atg gag  atg cag      6444
Lys Lys Lys  Gln Val Lys Leu Asp  Glu Gln Leu Met Glu  Met Gln
       1955                 1960                 1965

cac ctg agg  tcc act gcg acg cct  agc ccg tcc cct cat  gct tgg      6489
His Leu Arg  Ser Thr Ala Thr Pro  Ser Pro Ser Pro His  Ala Trp
       1970                 1975                 1980
```

```
gat ttg cag  ctg ctc cag cag caa  gcc tgt ccg atg gtg  ccc agg      6534
Asp Leu Gln  Leu Leu Gln Gln Gln  Ala Cys Pro Met Val  Pro Arg
        1985                 1990                 1995

gag cag ttt  ctg cag ctt caa cgc  cag ctg ctg cag gca  gaa agg      6579
Glu Gln Phe  Leu Gln Leu Gln Arg  Gln Leu Leu Gln Ala  Glu Arg
        2000                 2005                 2010

ata aac cag  cac ctg cag gag gaa  ctt gaa aac agg acc  tcc gaa      6624
Ile Asn Gln  His Leu Gln Glu Glu  Leu Glu Asn Arg Thr  Ser Glu
        2015                 2020                 2025

acc aac aca  cca cag gga aac cag  gaa caa ctg gta act  gtc atg      6669
Thr Asn Thr  Pro Gln Gly Asn Gln  Glu Gln Leu Val Thr  Val Met
        2030                 2035                 2040

gag gaa cga  atg ata gaa gtt gaa  cag aaa ctg aaa cta  gtg aaa      6714
Glu Glu Arg  Met Ile Glu Val Glu  Gln Lys Leu Lys Leu  Val Lys
        2045                 2050                 2055

agg ctt ctt  caa gag aaa gtg aat  cag ctc aaa gaa caa  ctc tgc      6759
Arg Leu Leu  Gln Glu Lys Val Asn  Gln Leu Lys Glu Gln  Leu Cys
        2060                 2065                 2070

aag aac act  aag gca gac gca atg  gtg aag gac ttg tat  gtt gaa      6804
Lys Asn Thr  Lys Ala Asp Ala Met  Val Lys Asp Leu Tyr  Val Glu
        2075                 2080                 2085

aat gcc cag  ttg ttg aaa gct ctg  gaa gtg act gaa cag  cga cag      6849
Asn Ala Gln  Leu Leu Lys Ala Leu  Glu Val Thr Glu Gln  Arg Gln
        2090                 2095                 2100

aaa aca gca  gag aag aaa aat tac  ctc ctg gag gag aag  att gcc      6894
Lys Thr Ala  Glu Lys Lys Asn Tyr  Leu Leu Glu Glu Lys  Ile Ala
        2105                 2110                 2115

agc ctc agt  aat ata gtt agg aat  ctg aca cca gcg cca  ttg act      6939
Ser Leu Ser  Asn Ile Val Arg Asn  Leu Thr Pro Ala Pro  Leu Thr
        2120                 2125                 2130

tct aca cct  cct ttg agg tca tagccaaacc aaagggtaca ctcatatttg      6990
Ser Thr Pro  Pro Leu Arg Ser
        2135

tgcactttac tgaaatagat gaacatttca gtaggttctc aacttaaaat taagcctaac  7050

ctaaaactgc cagcaacaca actggagttt ccatttatca taattagttt ttctaaatag  7110

acccttatgg gagtttgaaa ataaatactc acatatttca ctacttaaat tattcccaag  7170

atttgaattt attttaaaat tttaatagcc accaagaatg tggacatatg aaaattcaag  7230

aacctaaaaa ataccagttt tgaatgagtt tttgtggttt tggtttttta attattacaa  7290

atctatgtgt aaaatctaga tatttgaagt ttgagatctg atgagaatgg ttgttataaa  7350

ctttatttta aaaccaaatt taggtgttct tacatattta aatactggaa agtcattata  7410

atagttttgg ttctttgaat tggtagacaa ttagtagagt ataattggtt aggaggcagg  7470

gcttattaag tggttattaa ccgctgacat cagacaaacc caaatctgta gaattctaac  7530

ctcctaacac ctgtgacagt attaccactc ttcttgtatt atagatttag aactgattta  7590
```

ctcaattgca ctcttaacta atgttaaaag cttacttgct ttaaacagcc ttttcttctt 7650

tctcttaaaa gtttcatttg gggagctggt cttctaagaa acggataaag ccacataatt 7710

aaagcagttg aactagaggg aaagcactga acaaaccact ttggagtaaa tagctactct 7770

tagaaaagag ggataagcag accatgtagg ttttctgtct ctcaaatctt agagttcata 7830

aatttacttg aggttgcctc aagaactcag ggaacaatac tgtaaactgt cttcctgaac 7890

tactgtaggg cctctctaag aatttgaaat gtataaacca tgtgacctca tttatttgtc 7950

ttatatattt acagccatac tagaattttt atttctacgt ttttagtaaa tttaatattc 8010

tgggggaaaa aaggccttga ttttagggtt aaaaacctga cttatagaag agtttattta 8070

atataggtca aaatttttctg tgtttcttat tccttctata cctcaaatct gattctaaga 8130

atttcttact gtgataatca ttggcatgcc acctgaggtc aaggagtgcc aaataggact 8190

ttccactcat gctcaagatc aaaactttat agaacagtca acattttaga ttcggtaacc 8250

ttttttttct tccaattata atctctgctt ctagccactt ccgccagcag ttggtggaag 8310

acttactagg tgcagggcac tttccaagtt catcacaaca acctgcttgt tttcatgaga 8370

caataatccg aaaagttcgc tttgatatat tcctggaggg ccaagcccat ctatttacaa 8430

aaggtgaaca gcaaaatcaa gcactgcttt atgggcagga acacaagaga aagcaaactg 8490

cccaagaagt catcatgtca gaaactcaat ctcaacaaaa taatttccat cagggaactt 8550
cagggtttct tgggggctta tgagtctcac cggtcaaccc aggaggcctc actacaagag 8610

ccttgacaag gcactgtttt ttgtgggact gggagttcac actgatgaag caaacctttg 8670

aatttttgca cagctcttgt cagaaagccc tgagttcccc ctggataaag agttaatttt 8730

aatccttccc tataattata cttcaaaata tttgacatct gctattatgc cttctttaga 8790

tctttcttct gcggtgcaga catttctagt aagtgtttga ctacttgtat ggcattagct 8850

ttcacagaaa attgtttcac ttaaaactgt ggattggcct aggctaagga caaaaataaa 8910

ctaagtacct gtagtgtatt tatgtgatat gtgtcaagtt actcaaagtt attgctgttg 8970

gaactgaaca ataatatttc ccagatagct ggccttagca tgtgatcacg gttgttgtat 9030

ttttaatttt tgtcttttac agtatgagag gtgtaggtta atttgtttat ttcctataaa 9090

tttgtattta tgtgtatata aaatgtacaa tgaatgtaaa tatgactttc tggaaagttt 9150

agactacatt tagaatctct attcaaaatc aaaatgctgc tcaaatgaat ttaaccaaca 9210

tctaggtgct taatttctca ttttatccca cttatgagat tgggaaaaag atcaatatga 9270

gaaataccat acagatacct taaatgtatg catttgtgca acaatttttg agaaggtgag 9330

tggcaattta taatttagtt ggcaatttat aatagaactt atagctttta aaagactttt 9390

taaagacatt aaatgtaaac ttaaaaatgt ttagatcttg tttcaaactt tacaatagca 9450

ttcttcaaaa tattaagtta tatattttat aggcatttag ttgcttatta aaagcactga 9510

ttttcaaact ttttgattta agaacaatta tttaagatcg tctcagaaga tgggatcttc 9570

gtttcaagaa aagggaatca agtttgcctt tgagataata cgttacacta agaaaaggaa 9630

aatgtggata gtaaaaccca cctctctcat cctattgtac tctcttctgc tttttagaag 9690

cctgcactta agcttagatt tgtgaaggga gagtagaagg ggagaagtag aaccacagtg 9750

ttttatttat ttttctaaaa ctcttactaa atccagattt tttaaactgt tttaaatgtg 9810

aattcttccc agaaatttca atgcattgca tatttagcct tcggcatatt tttcatgaat 9870

agatcatgaa gtcataggct tccaaggcat aggaagagat cttgcaggtc tagtatttta 9930

ataatgcact attacccagg gcagatatta tgagaaactg tttcttctct aagggtttat 9990

ggcagacttt gctttttaa catgtgagaa atgaattttt tattttgtga tttatgtgat 10050

ttcttttgct gagtgaagga aaggagaaat tgttgctatt gtcagcatct taaaggtatt 10110

tccagtcaag gcaaggctaa gtgctttgtg atagtattaa gcaagtcatg ttttgaatgg 10170

attacctgta gtgactcatt ggaatgatat aattatacaa gtaatgccaa aaaccaagtc 10230

aaagcctaat taaccaaagc actcatttaa aaatcatcat gtttggacct atctggacct 10290

ctcagcactg taaaatagtt ttggttttgt ggcatatgaa tagctgttta acaaatcaaa 10350

gttagctttt tgcttctcag ctttttggg caatacaagt taagttctta atggggagac 10410
attatcatgg catgacttaa gggaacattg gtttgtgaag gaaaaacaga ttatctaaag 10470

ccatctctat gtttctgttc agataaagat taatgagttc tgtgtttata tcagctttgt 10530

atatttcatc ttagccattc tatcctagaa agattttaat gtgagcttaa gatgtaaata 10590

aataattttg caaacatgaa aaaaaaaaaa aaaaa 10625

<210> 4
<211> 2139
<212> PRT
<213> Homo sapiens

<400> 4
Met Ala Glu Val Thr Val Pro Arg Val Tyr Val Val Phe Gly Ile His
1 5 10 15

Cys Ile Met Ala Lys Ala Ser Ser Asp Val Gln Val Ser Gly Phe His
20 25 30

Arg Lys Ile Gln His Val Lys Asn Glu Leu Cys His Met Leu Ser Leu
35 40 45

Glu Glu Val Ala Pro Val Leu Gln Gln Thr Leu Leu Gln Asp Asn Leu
50 55 60

Leu Gly Arg Val His Phe Asp Gln Phe Lys Glu Ala Leu Ile Leu Ile
65 70 75 80

Leu Ser Arg Thr Leu Ser Asn Glu Glu His Phe Gln Glu Pro Asp Cys
85 90 95

Ser Leu Glu Ala Gln Pro Lys Tyr Val Arg Gly Gly Lys Arg Tyr Gly
100 105 110

Arg Arg Ser Leu Pro Glu Phe Gln Glu Ser Val Glu Glu Phe Pro Glu

```
        115                 120                 125

Val Thr Val Ile Glu Pro Leu Asp Glu Glu Ala Arg Pro Ser His Ile
    130                 135                 140

Pro Ala Gly Asp Cys Ser Glu His Trp Lys Thr Gln Arg Ser Glu Glu
145                 150                 155                 160

Tyr Glu Ala Glu Gly Gln Leu Arg Phe Trp Asn Pro Asp Asp Leu Asn
                165                 170                 175

Ala Ser Gln Ser Gly Ser Ser Pro Pro Gln Asp Trp Ile Glu Glu Lys
            180                 185                 190

Leu Gln Glu Val Cys Glu Asp Leu Gly Ile Thr Arg Asp Gly His Leu
        195                 200                 205

Asn Arg Lys Lys Leu Val Ser Ile Cys Glu Gln Tyr Gly Leu Gln Asn
    210                 215                 220

Val Asp Gly Glu Met Leu Glu Glu Val Phe His Asn Leu Asp Pro Asp
225                 230                 235                 240

Gly Thr Met Ser Val Glu Asp Phe Phe Tyr Gly Leu Phe Lys Asn Gly
                245                 250                 255
Lys Ser Leu Thr Pro Ser Ala Ser Thr Pro Tyr Arg Gln Leu Lys Arg
            260                 265                 270

His Leu Ser Met Gln Ser Phe Asp Glu Ser Gly Arg Arg Thr Thr Thr
        275                 280                 285

Ser Ser Ala Met Thr Ser Thr Ile Gly Phe Arg Val Phe Ser Cys Leu
    290                 295                 300

Asp Asp Gly Met Gly His Ala Ser Val Glu Arg Ile Leu Asp Thr Trp
305                 310                 315                 320

Gln Glu Glu Gly Ile Glu Asn Ser Gln Glu Ile Leu Lys Ala Leu Asp
                325                 330                 335

Phe Ser Leu Asp Gly Asn Ile Asn Leu Thr Glu Leu Thr Leu Ala Leu
            340                 345                 350

Glu Asn Glu Leu Leu Val Thr Lys Asn Ser Ile His Gln Ala Ala Leu
        355                 360                 365

Ala Ser Phe Lys Ala Glu Ile Arg His Leu Leu Glu Arg Val Asp Gln
    370                 375                 380

Val Val Arg Glu Lys Glu Lys Leu Arg Ser Asp Leu Asp Lys Ala Glu
385                 390                 395                 400

Lys Leu Lys Ser Leu Met Ala Ser Glu Val Asp Asp His His Ala Ala
                405                 410                 415

Ile Glu Arg Arg Asn Glu Tyr Asn Leu Arg Lys Leu Asp Gly Glu Tyr
            420                 425                 430

Lys Glu Arg Ile Ala Ala Leu Lys Asn Glu Leu Arg Lys Glu Arg Glu
        435                 440                 445

Gln Ile Leu Gln Gln Ala Gly Lys Gln Arg Leu Glu Leu Glu Gln Glu
    450                 455                 460
```

Ile Glu Lys Ala Lys Thr Glu Glu Asn Tyr Ile Arg Asp Arg Leu Ala
465 470 475 480

Leu Ser Leu Lys Glu Asn Ser Arg Leu Glu Asn Glu Leu Leu Glu Asn
485 490 495

Ala Glu Lys Leu Ala Glu Tyr Glu Asn Leu Thr Asn Lys Leu Gln Arg
500 505 510

Asn Leu Glu Asn Val Leu Ala Glu Lys Phe Gly Asp Leu Asp Pro Ser
515 520 525

Ser Ala Glu Phe Phe Leu Gln Glu Glu Arg Leu Thr Gln Met Arg Asn
530 535 540

Glu Tyr Glu Arg Gln Cys Arg Val Leu Gln Asp Gln Val Asp Glu Leu
545 550 555 560

Gln Ser Glu Leu Glu Glu Tyr Arg Ala Gln Gly Arg Val Leu Arg Leu
565 570 575

Pro Leu Lys Asn Ser Pro Ser Glu Glu Val Glu Ala Asn Ser Gly Gly
580 585 590

Ile Glu Pro Glu His Gly Leu Gly Ser Glu Glu Cys Asn Pro Leu Asn
595 600 605

Met Ser Ile Glu Ala Glu Leu Val Ile Glu Gln Met Lys Glu Gln His
610 615 620

His Arg Asp Ile Cys Cys Leu Arg Leu Glu Leu Glu Asp Lys Val Arg
625 630 635 640

His Tyr Glu Lys Gln Leu Asp Glu Thr Val Val Ser Cys Lys Lys Ala
645 650 655

Gln Glu Asn Met Lys Gln Arg His Glu Asn Glu Thr Arg Thr Leu Glu
660 665 670

Lys Gln Ile Ser Asp Leu Lys Asn Glu Ile Ala Glu Leu Gln Gly Gln
675 680 685

Ala Ala Val Leu Lys Glu Ala His His Glu Ala Thr Cys Arg His Glu
690 695 700

Glu Glu Lys Lys Gln Leu Gln Val Lys Leu Glu Glu Glu Lys Thr His
705 710 715 720

Leu Gln Glu Lys Leu Arg Leu Gln His Glu Met Glu Leu Lys Ala Arg
725 730 735

Leu Thr Gln Ala Gln Ala Ser Phe Glu Arg Glu Arg Glu Gly Leu Gln
740 745 750

Ser Ser Ala Trp Thr Glu Glu Lys Val Arg Gly Leu Thr Gln Glu Leu
755 760 765

Glu Gln Phe His Gln Glu Gln Leu Thr Ser Leu Val Glu Lys His Thr
770 775 780

Leu Glu Lys Glu Glu Leu Arg Lys Glu Leu Leu Glu Lys His Gln Arg
785 790 795 800

Glu Leu Gln Glu Gly Arg Glu Lys Met Glu Thr Glu Cys Asn Arg Arg

805 810 815

Thr Ser Gln Ile Glu Ala Gln Phe Gln Ser Asp Cys Gln Lys Val Thr
820 825 830

Glu Arg Cys Glu Ser Ala Leu Gln Ser Leu Glu Gly Arg Tyr Arg Gln
835 840 845

Glu Leu Lys Asp Leu Gln Glu Gln Gln Arg Glu Glu Lys Ser Gln Trp
850 855 860

Glu Phe Glu Lys Asp Glu Leu Thr Gln Glu Cys Ala Glu Ala Gln Glu
865 870 875 880

Leu Leu Lys Glu Thr Leu Lys Arg Glu Lys Thr Thr Ser Leu Val Leu
885 890 895

Thr Gln Glu Arg Glu Met Leu Glu Lys Thr Tyr Lys Glu His Leu Asn
900 905 910
Ser Met Val Val Glu Arg Gln Gln Leu Leu Gln Asp Leu Glu Asp Leu
915 920 925

Arg Asn Val Ser Glu Thr Gln Gln Ser Leu Leu Ser Asp Gln Ile Leu
930 935 940

Glu Leu Lys Ser Ser His Lys Arg Glu Leu Arg Glu Arg Glu Glu Val
945 950 955 960

Leu Cys Gln Ala Gly Ala Ser Glu Gln Leu Ala Ser Gln Arg Leu Glu
965 970 975

Arg Leu Glu Met Glu His Asp Gln Glu Arg Gln Glu Met Met Ser Lys
980 985 990

Leu Leu Ala Met Glu Asn Ile His Lys Ala Thr Cys Glu Thr Ala Asp
995 1000 1005

Arg Glu Arg Ala Glu Met Ser Thr Glu Ile Ser Arg Leu Gln Ser
1010 1015 1020

Lys Ile Lys Glu Met Gln Gln Ala Thr Ser Pro Leu Ser Met Leu
1025 1030 1035

Gln Ser Gly Cys Gln Val Ile Gly Glu Glu Glu Val Glu Gly Asp
1040 1045 1050

Gly Ala Leu Ser Leu Leu Gln Gln Gly Glu Gln Leu Leu Glu Glu
1055 1060 1065

Asn Gly Asp Val Leu Leu Ser Leu Gln Arg Ala His Glu Gln Ala
1070 1075 1080

Val Lys Glu Asn Val Lys Met Ala Thr Glu Ile Ser Arg Leu Gln
1085 1090 1095

Gln Arg Leu Gln Lys Leu Glu Pro Gly Leu Val Met Ser Ser Cys
1100 1105 1110

Leu Asp Glu Pro Ala Thr Glu Phe Phe Gly Asn Thr Ala Glu Gln
1115 1120 1125

Thr Glu Gln Phe Leu Gln Gln Asn Arg Thr Lys Gln Val Glu Gly
1130 1135 1140

Val Thr Arg Arg His Val Leu Ser Asp Leu Glu Asp Asp Glu Val
1145 1150 1155

Arg Asp Leu Gly Ser Thr Gly Thr Ser Ser Val Gln Arg Gln Glu
1160 1165 1170

Val Lys Ile Glu Glu Ser Glu Ala Ser Val Glu Gly Phe Ser Glu
1175 1180 1185

Leu Glu Asn Ser Glu Glu Thr Arg Thr Glu Ser Trp Glu Leu Lys
1190 1195 1200

Asn Gln Ile Ser Gln Leu Gln Glu Gln Leu Met Met Leu Cys Ala
1205 1210 1215

Asp Cys Asp Arg Ala Ser Glu Lys Lys Gln Asp Leu Leu Phe Asp
1220 1225 1230

Val Ser Val Leu Lys Lys Lys Leu Lys Met Leu Glu Arg Ile Pro
1235 1240 1245

Glu Ala Ser Pro Lys Tyr Lys Leu Leu Tyr Glu Asp Val Ser Arg
1250 1255 1260

Glu Asn Asp Cys Leu Gln Glu Glu Leu Arg Met Met Glu Thr Arg
1265 1270 1275

Tyr Asp Glu Ala Leu Glu Asn Asn Lys Glu Leu Thr Ala Glu Val
1280 1285 1290

Phe Arg Leu Gln Asp Glu Leu Lys Lys Met Glu Glu Val Thr Glu
1295 1300 1305

Thr Phe Leu Ser Leu Glu Lys Ser Tyr Asp Glu Val Lys Ile Glu
1310 1315 1320

Asn Glu Gly Leu Asn Val Leu Val Leu Arg Leu Gln Gly Lys Ile
1325 1330 1335

Glu Lys Leu Gln Glu Ser Val Val Gln Arg Cys Asp Cys Cys Leu
1340 1345 1350

Trp Glu Ala Ser Leu Glu Asn Leu Glu Ile Glu Pro Asp Gly Asn
1355 1360 1365

Ile Leu Gln Leu Asn Gln Thr Leu Glu Glu Cys Val Pro Arg Val
1370 1375 1380

Arg Ser Val His His Val Ile Glu Glu Cys Lys Gln Glu Asn Gln
1385 1390 1395

Tyr Leu Glu Gly Asn Thr Gln Leu Leu Glu Lys Val Lys Ala His
1400 1405 1410

Glu Ile Ala Trp Leu His Gly Thr Ile Gln Thr His Gln Glu Arg
1415 1420 1425

Pro Arg Val Gln Asn Gln Val Ile Leu Glu Glu Asn Thr Thr Leu
1430 1435 1440

Leu Gly Phe Gln Asp Lys His Phe Gln His Gln Ala Thr Ile Ala
1445 1450 1455

Glu Leu Glu Leu Glu Lys Thr Lys Leu Gln Glu Leu Thr Arg Lys

1460 1465 1470

Leu Lys Glu Arg Val Thr Ile Leu Val Lys Gln Lys Asp Val Leu
1475 1480 1485

Ser His Gly Glu Lys Glu Glu Glu Leu Lys Ala Met Met His Asp
1490 1495 1500

Leu Gln Ile Thr Cys Ser Glu Met Gln Gln Lys Val Glu Leu Leu
1505 1510 1515

Arg Tyr Glu Ser Glu Lys Leu Gln Gln Glu Asn Ser Ile Leu Arg
1520 1525 1530
Asn Glu Ile Thr Thr Leu Asn Glu Glu Asp Ser Ile Ser Asn Leu
1535 1540 1545

Lys Leu Gly Thr Leu Asn Gly Ser Gln Glu Glu Met Trp Gln Lys
1550 1555 1560

Thr Glu Thr Val Lys Gln Glu Asn Ala Ala Val Gln Lys Met Val
1565 1570 1575

Glu Asn Leu Lys Lys Gln Ile Ser Glu Leu Lys Ile Lys Asn Gln
1580 1585 1590

Gln Leu Asp Leu Glu Asn Thr Glu Leu Ser Gln Lys Asn Ser Gln
1595 1600 1605

Asn Gln Glu Lys Leu Gln Glu Leu Asn Gln Arg Leu Thr Glu Met
1610 1615 1620

Leu Cys Gln Lys Glu Lys Glu Pro Gly Asn Ser Ala Leu Glu Glu
1625 1630 1635

Arg Glu Gln Glu Lys Phe Asn Leu Lys Glu Glu Leu Glu Arg Cys
1640 1645 1650

Lys Val Gln Ser Ser Thr Leu Val Ser Ser Leu Glu Ala Glu Leu
1655 1660 1665

Ser Glu Val Lys Ile Gln Thr His Ile Val Gln Gln Glu Asn His
1670 1675 1680

Leu Leu Lys Asp Glu Leu Glu Lys Met Lys Gln Leu His Arg Cys
1685 1690 1695

Pro Asp Leu Ser Asp Phe Gln Gln Lys Ile Ser Ser Val Leu Ser
1700 1705 1710

Tyr Asn Glu Lys Leu Leu Lys Glu Lys Glu Ala Leu Ser Glu Glu
1715 1720 1725

Leu Asn Ser Cys Val Asp Lys Leu Ala Lys Ser Ser Leu Leu Glu
1730 1735 1740

His Arg Ile Ala Thr Met Lys Gln Glu Gln Lys Ser Trp Glu His
1745 1750 1755

Gln Ser Ala Ser Leu Lys Ser Gln Leu Val Ala Ser Gln Glu Lys
1760 1765 1770

Val Gln Asn Leu Glu Asp Thr Val Gln Asn Val Asn Leu Gln Met
1775 1780 1785

Ser Arg Met Lys Ser Asp Leu Arg Val Thr Gln Gln Glu Lys Glu
1790 1795 1800

Ala Leu Lys Gln Glu Val Met Ser Leu His Lys Gln Leu Gln Asn
1805 1810 1815

Ala Gly Gly Lys Ser Trp Ala Pro Glu Ile Ala Thr His Pro Ser
1820 1825 1830

Gly Leu His Asn Gln Gln Lys Arg Leu Ser Trp Asp Lys Leu Asp
1835 1840 1845

His Leu Met Asn Glu Glu Gln Gln Leu Leu Trp Gln Glu Asn Glu
1850 1855 1860

Arg Leu Gln Thr Met Val Gln Asn Thr Lys Ala Glu Leu Thr His
1865 1870 1875

Ser Arg Glu Lys Val Arg Gln Leu Glu Ser Asn Leu Leu Pro Lys
1880 1885 1890

His Gln Lys His Leu Asn Pro Ser Gly Thr Met Asn Pro Thr Glu
1895 1900 1905

Gln Glu Lys Leu Ser Leu Lys Arg Glu Cys Asp Gln Phe Gln Lys
1910 1915 1920

Glu Gln Ser Pro Ala Asn Arg Lys Val Ser Gln Met Asn Ser Leu
1925 1930 1935

Glu Gln Glu Leu Glu Thr Ile His Leu Glu Asn Glu Gly Leu Lys
1940 1945 1950

Lys Lys Gln Val Lys Leu Asp Glu Gln Leu Met Glu Met Gln His
1955 1960 1965

Leu Arg Ser Thr Ala Thr Pro Ser Pro Ser Pro His Ala Trp Asp
1970 1975 1980

Leu Gln Leu Leu Gln Gln Gln Ala Cys Pro Met Val Pro Arg Glu
1985 1990 1995

Gln Phe Leu Gln Leu Gln Arg Gln Leu Leu Gln Ala Glu Arg Ile
2000 2005 2010

Asn Gln His Leu Gln Glu Glu Leu Glu Asn Arg Thr Ser Glu Thr
2015 2020 2025

Asn Thr Pro Gln Gly Asn Gln Glu Gln Leu Val Thr Val Met Glu
2030 2035 2040

Glu Arg Met Ile Glu Val Glu Gln Lys Leu Lys Leu Val Lys Arg
2045 2050 2055

Leu Leu Gln Glu Lys Val Asn Gln Leu Lys Glu Gln Leu Cys Lys
2060 2065 2070

Asn Thr Lys Ala Asp Ala Met Val Lys Asp Leu Tyr Val Glu Asn
2075 2080 2085

Ala Gln Leu Leu Lys Ala Leu Glu Val Thr Glu Gln Arg Gln Lys
2090 2095 2100

Thr Ala Glu Lys Lys Asn Tyr Leu Leu Glu Glu Lys Ile Ala Ser

2105 2110 2115

Leu Ser Asn Ile Val Arg Asn Leu Thr Pro Ala Pro Leu Thr Ser
2120 2125 2130

Thr Pro Pro Leu Arg Ser
2135